COLECȚIA „TEORIA PROBABILITĂȚILOR" IV

GH. MIHOC și V. URSEANU

MATEMATICI APLICATE ÎN STATISTICĂ

EDITURA ACADEMIEI REPUBLICII POPULARE ROMÎNE

Scanned with OKEN Scanner

COLECȚIA „TEORIA PROBABILITĂȚILOR" IV

GH. MIHOC și V. URSEANU

MATEMATICI APLICATE ÎN STATISTICĂ

EDITURA ACADEMIEI REPUBLICII POPULARE ROMÎNE
1962

Scanned with OKEN Scanner

PREFAȚĂ

În ultimii ani au apărut în matematică teorii noi, care prezintă o importanță considerabilă în aplicațiile practice. Programarea liniară cu multiplele ei extinderi și programarea dinamică permit întocmirea unor programe optime în diferite domenii de activitate. Progresul acestor teorii a fost posibil datorită apariției mașinilor electronice de calcul, care au permis transformarea efectivă a teoriei în practică. Calculul cu numere întîmplătoare, cunoscut și sub numele impropriu de metoda Monte Carlo, a dat și el un impuls dezvoltării acestor teorii, întrucît oferă un instrument de calcul natural pentru elementele numerice care intervin în programarea stohastică. În calculul indicilor, care ocupă un loc de frunte în statistica din țările socialiste, s-au elaborat de curînd noi teorii, bazate pe o analiză matematică mai profundă, datorite statisticienilor sovietici. Practica a pus în aplicarea metodei selecției probleme noi, a căror rezolvare face apel la cunoștințe avansate în domeniul matematicii.

Am socotit util să oferim statisticienilor din țara noastră o expunere sistematică a acestor teorii recente, care nu se găsesc în literatura noastră de specialitate.

Dificultatea principală care a stat în fața noastră a fost că, plecînd de la faptul că aceste teorii interesează în egală măsură pe matematicieni și statisticienii economiști, adică două categorii de cititori cu pregătire științifică diferită, trebuia să scriem o carte de matematică, accesibilă pînă la un punct și pentru nespecialiști în domeniul matematicii. În acest scop, teoriile și demonstrațiile, care interesează numai pe matematicieni, au fost tipărite cu litere cu caracter mic. Am scris o parte de calculul probabilităților și alta de distribuții statistice, reduse la elemente esențiale, fără a sacrifica exactitatea științifică. Ele servesc ca bază teoretică pentru celelalte capitole. Am dat numeroase exemple numerice, care ușurează înțelegerea teoriilor generale.

Colegii noștri V. Trebici, G. Theiler, L. Tövissi, A. Sălăgeanu, și I. Stoichiță, citind total sau parțial manuscrisul, au făcut observații interesante, de care am ținut seamă la redactare. L. Tövissi și N. Rancu ne-au oferit exemple din practica lor personală în aplicarea statisticii matematice în industrie, pe care le-am folosit în carte. Tuturor le mulțumim și pe această cale.

AUTORII

Scanned with OKEN Scanner

TABLA DE MATERII

Pag.

Partea întîi

CALCULUL PROBABILITĂŢILOR

Capitolul I

Fundamentarea teoretică a calculului probabilităţilor. Noţiunea de probabilitate 13

1. Fundamentarea teoretică a calculului probabilităţilor 13
2. Probabilitate, frecvenţă 15
3. Schema lui Kolmogorov 23

Capitolul II

Determinarea probabilităţilor. Reguli de calcul pentru evenimente în număr finit 28

1. Consideraţii generale 28
2. Reguli de calcul 31

Capitolul III

Scheme clasice din teoria probabilităţilor 46

1. Schema urnei Bernoulli 46
2. Schema multinomială 48
3. Schema bilei nerepetate 51
4. Schema lui Poisson 52
5. Inegalitatea lui Boole 54

Capitolul IV

Valorii medii 56

1. Variabilă aleatoare. Lege de probabilitate. Valoare medie (Speranţă matematică) 56
2. Valoarea medie a sumei şi a produsului a două variabile aleatoare 59

Pag.

3. Valoarea medie a produsului variabilelor independente dintr-un sistem 60
4. Valoarea medie a numărului de bile albe obţinut în n extracţii succesive dintr-o urnă Bernoulli . 61
5. Momente. Media armonică. Media geometrică 63
6. Valoarea medie pătratică a unei sume de variabile independente dintr-un sistem 65
7. Inegalităţi între valorile medii 66
8. Consideraţii generale asupra valorilor medii ale variabilelor aleatoare. Semnificaţia valorilor medii 68

Capitolul V

Dispersie. Abateri medii 72

1. Consideraţii generale 72
2. Abatere medie 73
3. Dispersie. Abaterea medie pătratică 74
4. Abaterea medie pătratică a numărului de bile albe obţinute în n extracţii în cazul urnei Bernoulli 75
5. Abaterea medie pătratică a unei sume de variabile aleatoare . . . 76
6. Momente centrate 79
7. Inegalitatea lui Cebîşev 79

Capitolul VI

Funcţie de repartiţie. Funcţie caracteristică 85

1. Funcţie de repartiţie. Densitate de probabilitate 85
2. Funcţie caracteristică 95

Scanned with OKEN Scanner

Pag.

Capitolul VII

Legi de probabilitate clasice 97

1. Legea binomială 97
2. Legea normală 105
3. Lege polinomială (multinomială) 126
4. Legea hipergeometrică 128
5. Legea Poisson (legea evenimentelor rare) 132
6. Legea de probabilitate χ^2 135
7. Legea de probabilitate Student . . 145

Capitolul VIII

Legea numerelor mari 149

1. Considerații generale 149
2. Teorema lui Bernoulli 151
3. Teorema lui Poisson 152
4. Teorema lui Cantelli 152
5. Un exemplu de convergență sigură 154
6. Convergența variabilei aleatoare în general 155
7. Esența legii numerelor mari. Câteva observații 156

Partea a doua

ELEMENTE DE ANALIZĂ A DISTRIBUȚIILOR STATISTICE

Capitolul IX

Distribuții statistice cu o singură caracteristică 161

1. Noțiuni introductive. Caracteristici variabile 161
2. Funcție teoretică. Funcție experimentală 164
3. Diferite tipuri de distribuții statistice și elementele lor generale 165
4. Densitatea distribuției 174
5. Frecvența cumulată — integrală a densității distribuției 177
6. Descompunerea intervalului și determinarea frecvențelor parțiale corespunzătoare 179
7. Reprezentarea grafică a distribuțiilor cu o singură caracteristică 185
8. Forma distribuțiilor de variație . 193

Capitolul X

Mărimi medii folosite în practica statisticii 199

1. Considerații generale 199
2. Media aritmetică și proprietățile ei 200
3. Media armonică 208
4. Media pătratică 212

Pag.

5. Media cubică 213
6. Media geometrică 214
7. Medii de ordin superior. Momente 216
8. Momente inițiale. Momente obișnuite. Momente centrate. 217
9. Formule de trecere de la momente obișnuite la momente centrate. Calculul mediei aritmetice prin metoda momentelor 219
10. Medii de structură 227
11. Indicatorii variației. Asimetrie . . 236

Capitolul XI

Distribuții multidimensionale (cu mai multe caracteristici) 245

1. Considerații generale 245
2. Elemente de analiză a distribuțiilor multidimensionale 247
3. Valori medii condiționate. Funcție de regresie 256
4. Analiza legăturii statistice dintre caracteristicile distribuției. Metoda corelației 257
5. Drepte de regresie 266
6. Coeficientul de corelație 267
7. Corelația rangurilor 272
8. Distribuții cu trei dimensiuni . . 277
9. Problema regresiei 281
10. Corelație parțială și multiplă . . 282
11. Distribuții cu n dimensiuni . . . 287

Partea a treia

ANALIZA SERIILOR ÎN TIMP (DINAMICE)

Capitolul XII

Probleme generale ale elaborării și analizei seriilor dinamice . . . 291

1. Oscilațiile în timp ale variabilelor și construirea seriilor dinamice . . 291
2. Clasificarea seriilor dinamice și reprezentarea lor grafică 295
3. Indicatorii seriei dinamice 299

Capitolul XIII

Ajustarea și interpolarea seriilor dinamice 305

1. Procedee elementare de ajustare . 305
2. Metoda mediilor mobile 309
3. Metoda ajustării analitice. Alegerea curbei de tendință 315
4. Ajustarea cu o linie dreaptă prin metoda celor mai mici pătrate . 316
5. Alte metode de ajustare 321
6. Câteva considerații generale asupra ajustării. Formule de interpolare . 332

Scanned with OKEN Scanner

Pag.

Capitolul XIV

Metode de calcul al ritmului mediu de dezvoltare 339

1. Generalităţi 339
2. Metoda mediei geometrice 341
3. Metoda mediei parabolice 342
4. Metoda lui Froda pentru calculul ritmului mediu de creştere . . . 345
5. Metoda tendinţei exponenţiale . 354
6. Cîteva concluzii asupra procedeelor de calcul al ritmului mediu de creştere 355

Partea a patra

METODA INDICILOR

Capitolul XV

Esenţa indicilor. Clasificarea lor şi metodele de calcul 361

1. Noţiuni generale 361
2. Indici în formă de bază şi în formă de lanţ 363
3. Clasificarea indicilor după funcţiile lor cognitive. Schema lui V. Peregudov 366
4. Aplicaţii ale calculului indicilor. Indici ai preţului de cost 379
5. Analiza indicilor cu structură variabilă 406
6. Analiza indicelui caracteristicii funcţionale în ipoteza unei producţii eterogene 419

Capitolul XVI

Indici teritoriali 423

1. Consideraţii introductive 423
2. Alegerea bazei de comparaţie (raportare) şi a sistemului de ponderi 424
3. Metode de calcul 426
4. Observaţii şi concluzii 445

Capitolul XVII

Aplicaţii ale coeficientului de corelaţie liniară în lucrări de analiză prin metoda indicilor 457

1. Consideraţii introductive 457
2. Corelaţia dintre două mărimi medii calculate pe baza ponderilor diferite 458
3. Corelaţia dintre indicii agregaţi cu ponderi diferite. Indici ai preţurilor 461

Pag.

4. Corelaţia dintre indici ai mărimilor medii (de tip $C\left(I_{\alpha}=\frac{\bar{\alpha}_1}{\bar{\alpha}_0}\right)$) şi indici de alt tip 467

Capitolul XVIII

Aplicarea legii numerelor mari la elaborarea indicilor 475

1. Consideraţii introductive 475
2. Sistemul reprezentativ 478

Partea a cincea

METODA SELECŢIEI

Capitolul XIX

Metoda selecţiei şi principiul ei. Importanţa aplicării ei în cercetările statistice 517

1. Consideraţii generale. Selecţia şi principiul ei 517
2. Reprezentativitatea selecţiei. . . . 520
3. Cercetarea selectivă în lucrările de recensăminte şi anchete 523
4. Formularea matematică a principiului selecţiei 525

Capitolul XX

Structura diferitelor tipuri de selecţie. Sondaje cu o fază 531

1. Consideraţii generale 531
2. Sondajul simplu nerepetat . . . 532
3. Sondajul simplu repetat 554
4. Indicatorii erorilor relative ale sondajului 556
5. Cîteva observaţii generale asupra sondajului simplu 558
6. Sondajul mecanic 564
7. Sondajul tipic (stratificat) simplu . 565
8. Sondajul tipic proporţional sau reprezentativ 573
9. Sondajul tipic optim 576
10. Cîteva observaţii generale asupra sondajului tipic 584

Capitolul XXI

Sondaje în două sau mai multe faze şi alte sondaje 587

1. Sondajul în două faze, generalităţi 587
2. Sondajul în două faze cu o singură caracteristică 588
3. Sondajul în două faze cu mai multe caracteristici 594

Scanned with OKEN Scanner

Pag.

Capitolul XXII

Sondajul cu volum redus. Concluzii asupra metodei selecției 598

1. Considerații generale 598
2. Verificarea ipotezelor 599
3. Distribuția χ^2 600
4. Distribuția lui Student 608
5. Distribuția lui Kolmogorov . . . 613
6. Extinderea rezultatelor cercetării selecției asupra colectivității generale 623
7. Cîteva observații și concluzii generale asupra metodei selecției . . . 624

Partea a șasea

METODE MODERNE DE CALCUL STATISTIC

Capitolul XXIII

Metode statistice de calcul numeric. Calculul cu numere întîmplătoare . 629

1. Considerații generale 629
2. Calculul valorii lui π 630
3. Problema acului 632
4. Folosirea numerelor întîmplătoare 636
5. Calculul unei arii mărginite de o curbă convexă 641
6. Calculul unei integrale definite . 641
7. Mersul la întîmplare 644
8. Rătăcirea întîmplătoare uniformă pe o suprafață plană 650
9. Ecuația lui Laplace 651

Pag.

10. Aproximarea prin polinoame . . 652
11. Inversa unei matrice 654
12. Procese cu legături complete . . 656

Capitolul XXIV

Programarea liniară 659

1. Considerații generale 659
2. Metoda Simplex 665
3. Bazele matematice ale programării liniare 669
4. Problema transporturilor 681

Capitolul XXV

Balanța legăturilor dintre ramuri . 698

1. Considerații generale 698
2. Fundamentarea teoretică a balanței legăturilor dintre ramuri 699
3. Schema balanței legăturilor dintre ramuri 701
4. Coeficienții tehnici și unele aplicații ale balanței legăturilor dintre ramuri 704

Capitolul XXVI

Diverse generalizări ale metodei programării liniare 710

1. Caracterul statistic al programării liniare 710
2. Programare stohastică 715
3. Programarea dinamică 721

Tabela-anexă I 727
Tabela-anexă II 730
Tabela-anexă III 731

Scanned with OKEN Scanner

PARTEA ÎNTÎI

CALCULUL PROBABILITĂŢILOR

Scanned with OKEN Scanner

Scanned with OKEN Scanner

Capitolul I

Fundamentarea teoretică a calculului probabilităţilor. Noţiunea de probabilitate

1. Fundamentarea teoretică a calculului probabilităţilor

Calculul probabilităţilor constituie în prezent una dintre cele mai moderne ramuri ale matematicii contemporane. Ca şi geometria, mecanica raţională, algebra etc., calculul probabilităţilor s-a închegat într-o disciplină matematică riguros construită, capabilă să satisfacă cerinţele ştiinţelor naturii şi ale tehnicii moderne a căror dezvoltare cunoaşte astăzi un avînt fără precedent.

În procesul progresiv al cunoaşterii ştiinţifice, noţiunile de *probabilitate* şi *calculul probabilităţilor* şi-au lărgit şi îmbogăţit conţinutul. De la interpretarea lor clasică, încercată încă în secolul al XVII-lea pentru prima oară de cunoscuţii matematicieni P. Fermat *), B. Pascal **) şi J. Bernoulli ***), interpretare formulată, cum era şi firesc, la nivelul cunoş-

*) P. Fermat (1601—1665), mare matematician francez, deşi prin profesiunea de bază era jurist. P. Fermat este unul din creatorii teoriei numerelor. În domeniul geometriei, înainte de R. Descartes şi într-o formă mai sistematică, a dezvoltat teoria coordonatelor; P. Fermat s-a ocupat de probleme variate din domeniul matematicii. Opera lui a avut o influenţă considerabilă asupra dezvoltării continue a matematicii.

În fizică, de numele lui Fermat este legată formularea principiului fundamental al opticii geometrice.

**) B. Pascal (1623—1662), un ilustru matematician, fizician şi filozof francez. De mic copil Pascal a dovedit un talent deosebit pentru matematici, descoperind singur, la vîrsta de 12 ani, primele propoziţii ale geometriei lui Euclid. În 1639 (la 16 ani) a publicat o lucrare asupra conicelor în care a formulat una din teoremele fundamentale ale geometriei proiective. Preocupările lui în domeniul matematicii au fost multilaterale. În 1641 (după unii în 1642) el a conceput, cel dintîi, planul unei maşini de calcul.

În fizică, Pascal s-a ocupat cu studiul presiunii barometrice şi cu problemele de hidrostatică. Lucrările lui în teoria presiunii atmosferice, au avut o importanţă deosebită pentru meteorologie.

***) J. Bernoulli (1654—1705) a fost profesor de matematici la Universitatea din Basel, începînd din 1687. În colaborare cu fratele său mai mic I. Bernoulli, matematician şi el, a pus bazele calculului variaţional J. Bernoulli a rezolvat unele probleme ale analizei combinatorii şi în legătură cu studiul sumei de forma $1^m + 2^m + 3^m + \ldots n^m$ a descoperit numere numite mai tîrziu numerele lui Bernoulli. În domeniul calculului probabilităţilor J. Bernoulli a formulat o cunoscută teoremă, care-i poartă numele, un caz particular important al legii numerelor mari. J. Bernoulli s-a ocupat şi cu problemele fizicii.

tinţelor matematice din acel timp, pînă la interpretarea lor teoretică justă, în lumina materialismului dialectic, noţiunile de probabilitate şi calculul probabilităţilor au cunoscut o continuă dezvoltare, adîncindu-se, precizîndu-se şi ridicîndu-se la un nivel superior de abstracţie.

Trebuie să precizăm că abia în secolul nostru, odată cu dezvoltarea impetuoasă a ştiinţei şi tehnicii, noţiunile de bază ale teoriei probabilităţilor au căpătat un conţinut precis şi riguros ştiinţific. La acest lucru s-a ajuns datorită *construirii axiomatice a teoriei probabilităţilor*, concepută şi elaborată de ilustrul matematician sovietic A. N. Kolmogorov. Meritul acestei axiomatizări constă în faptul că *axiomele care stau la baza teoriei probabilităţilor rezultă din legile obiective ale fenomenelor reale de masă* şi sînt o *consecinţă* a lor. Totodată, axiomatizarea făcută de Kolmogorov leagă strîns teoria probabilităţilor de teoria funcţiilor de o variabilă reală, făcînd astfel din calculul probabilităţilor unul dintre cele mai moderne domenii ale matematicii.

În esenţă, în interpretarea materialistă, probabilitatea nu mai reprezintă doar o valoare teoretică, ruptă de realitate, obţinută prin calcule formale, ci rezultă din studiul adîncit al fenomenelor şi proceselor care au loc în natură şi în viaţa socială.

Este evident că întrucît la baza teoriei probabilităţilor stau anumite axiome, obţinute paralel cu dezvoltarea ştiinţelor naturii şi a ştiinţelor sociale, rezultatele teoriei probabilităţilor, obţinute pe bază de deducţii logice, se întorc la ştiinţele naturii şi chiar la unele domenii ale ştiinţelor sociale, pentru a găsi aici nenumărate aplicaţii.

Fundamentarea teoretică justă, materialist-dialectică, a calculului probabilităţilor şi-a trecut strălucit examenul în tot cursul dezvoltării impetuoase a ştiinţelor contemporane ale naturii.

Se confirmă astfel, odată mai mult, principiul ştiinţific al legăturii dialectice dintre teorie şi practică, principiul verificării în practică a tezelor teoretice, al dezvoltării teoriei pe baza generalizării experienţelor anterioare.

Vom conchide aşadar că *la baza calculului probabilităţilor stau tezele materialismului dialectic* a cărui metodă de cercetare ştiinţifică şi-a găsit o strălucită aplicare în cazul particular al schemei axiomatice elaborată de Kolmogorov.

Asupra esenţei acestei scheme ne vom opri pe scurt mai jos, după ce vom prezenta cîteva consideraţii generale şi precizări pe marginea definiţiei clasice a noţiunii de probabilitate. Acest lucru se impune cu atît mai mult cu cît axiomele lui Kolmogorov, în *anumite condiţii*, impuse de realitatea obiectivă a fenomenelor şi proceselor studiate, nu contrazic definiţia clasică a probabilităţilor. Ele ne arată cum anume putem folosi definiţia clasică. Aşadar, din punct de vedere matematic, cu noţiunea de probabilitate în sensul ei clasic se poate lucra. Este necesar însă ca de fiecare dată cercetătorul să se convingă că din punct de vedere practic sînt îndeplinite condiţiile la care ne-am referit mai sus, care justifică aplicarea schemelor clasice ale teoriei probabilităţilor.

Scanned with OKEN Scanner

2. Probabilitate, frecvență

Pentru lămurirea noțiunii de probabilitate vom considera cîteva exemple care ne vor conduce în mod natural la definirea ei.

Să luăm, de exemplu, o întreprindere care fabrică anumite produse standardizate și să admitem că din 100 de piese, 95 corespund standardului stabilit; în acest caz, putem afirma, folosind un termen des întrebuințat în vorbirea curentă, că luînd la întîmplare una din 100 de piese executate în condiții identice (nu intervin modificări importante în modul de funcționare a fabricii, în nivelul și calitatea tehnicii ei etc.), avem șansă mai mare de a scoate o piesă corespunzătoare standardului, deci o piesă bună, decît o piesă rebut. O asemenea afirmație se impune în mod rațional deoarece știm că din seria de 100 de piese fabricate, 95 din ele reprezintă *cazuri favorabile* de a se produce, la întîmplare, *evenimentul* scoaterii unei piese bune. Dimpotrivă, șansa de a scoate o piesă rebut este cu mult mai mică întrucît avem numai 5 asemenea piese, *cazuri nefavorabile* (pentru apariția unei piese bune). Desigur, nu în fiecare serie de 100 de piese se vor obține 95 de cazuri favorabile, într-o serie de 100 de piese pot fi favorabile mai multe cazuri, în alta pot fi mai puține decît 95. Dacă observăm însă un număr mare de serii de cîte 100 de piese, executate în condiții identice, vom constata că numărul mediu al pieselor corespunzătoare standardului stabilit este foarte apropiat de 95.

Numărul de 95% reprezintă pentru întreprinderea noastră un indicator al posibilităților ei. În măsura în care numărul seriilor de cîte 100 de piese devine mai mare, *stabilitatea* acestui indicator apare și mai evidentă. El ne permite să apreciem unele evenimente de *masă* care au avut loc și să tragem concluzii (previziuni) valabile pentru anumite operații de masă care vor avea loc în viitor. Importanța cunoașterii unui asemenea indicator este relevantă atît în ramura industriei a economiei naționale, unde furnizează date prețioase asupra calității producției, cît și în alte ramuri de activitate economică, în fizică, chimie, tehnică, astronomie, statistică social-economică (în acele domenii ale ei în care calculul probabilităților se aplică în mod rațional) etc.

În exemplul considerat mai sus, vom putea spune că, în medie, din fiecare 100 piese, 5 vor fi date la rebut. Folosind un termen matematic, care va fi precizat mai departe, vom spune că pentru o asemenea întreprindere, *probabilitatea* de a fabrica o piesă bună este de 95% sau 0,95 și, complementar, probabilitatea de a fabrica o piesă rebut, este de 5% sau 0,05.

În exemplul considerat mai observăm că, luînd *la întîmplare* o piesă dintr-o serie de 100 de piese, avem *posibilitatea* să obținem oricare din cele o sută piese; seria de 100 piese reprezintă numărul total de *cazuri posibile* și ele sînt *egal posibile* deoarece alegerea se face cu asigurarea deplină a caracterului ei întîmplător, piesele nediferind prin culoare sau greutate. Se vede clar că numărul cazurilor egal posibile este o sumă a cazurilor favorabile și nefavorabile.

Scanned with OKEN Scanner

Un exemplu de stabilitate statistică ni-l oferă și datele din tabelul 1.

Tabelul 1

Născuți vii pe sexe în orașul București, în anii 1948 și 1951–1958
(Proporția între băieți și fete)

Anii	Născuți vii:	
	sex masculin	sex feminin
1948	0,511	0,489
1951	0,514	0,486
1952	0,516	0,484
1953	0,514	0,486
1954	0,508	0,492
1955	0,520	0,480
1956	0,518	0,482
1957	0,515	0,485
1958	0,519	0,481

În sfîrșit, să mai considerăm un exemplu, de tipul celor de care ne vom folosi adeseori mai departe:

Avem două urne marcate prin literele A și B.

Urna A conține 100 de bile de aceeași formă și greutate din care 10 sînt de culoare albă, iar restul de alte culori. Urna B conține tot 100 de bile, din care însă 80 sînt albe.

Dorind să scoatem o bilă albă la întîmplare din fiecare urnă, vom afirma în mod intuitiv că există o probabilitate mai mare de a scoate o bilă albă din urna B decît din urna A.

Folosind și în acest exemplu termenii matematici precizați ca în primul exemplu, vom spune, referindu-ne la urna A, că avem în total 100 de cazuri posibile, fiindcă sînt 100 de bile în urnă dintre care putem scoate pe oricare și că avem 10 cazuri favorabile pentru a scoate o bilă albă la întîmplare și 90 de cazuri nefavorabile (în 90 de cazuri din 100, putem scoate o bilă de altă culoare decît cea albă); de asemenea, referindu-ne la urna B vom putea spune că și aici avem 100 de cazuri posibile din care însă 80 sînt cazuri favorabile pentru a scoate o bilă albă și 20 cazuri nefavorabile. Și la o urnă și la cealaltă, toate cazurile posibile sînt *egal* posibile, fiindcă nu le putem deosebi nevăzînd bilele.

Din ultimul exemplu considerat se poate observa și mai clar că dacă numărul cazurilor posibile este același, evenimentul de a fi scoasă la întîmplare o bilă albă are o șansă sau *probabilitate* mai mare de a se produce atunci cînd numărul cazurilor favorabile este mai mare. Dacă, invers, am presupune că în ambele urne avem 10 bile albe, dar că urna A are 100 de bile, iar B 150 de bile, atunci probabilitatea de a extrage o bilă albă din A este mai mare decît aceea de a extrage o asemenea bilă din urna B.

Rezultă, prin urmare, că dacă lăsăm neschimbat numărul cazurilor posibile iar numărul cazurilor favorabile crește, probabilitatea crește și ea.

Invers, dacă lăsăm neschimbat numărul cazurilor favorabile și facem să crească numărul cazurilor posibile, probabilitatea descrește.

Dacă și în ultimul exemplu — cu cele două ipoteze privind structura urnei — vom introduce un *indicator numeric* care să măsoare șansa de a se extrage o bilă albă, ținînd seama de modul de variație mai sus arătat al probabilității și amintindu-ne de proprietățile rapoartelor, este firesc ca acest indicator să fie ales ca fiind egal cu *raportul dintre numărul cazurilor favorabile și numărul cazurilor posibile*. În prima ipoteză probabilitatea de a extrage o bilă albă din urna A va fi egală deci cu : $\frac{10}{100} = \frac{1}{10} = 0{,}1$ sau 10%, iar din urna B ea va fi egală cu : $\frac{80}{100} = \frac{8}{10} = 0{,}8$ sau 8%; în ipoteza a doua, în cazul cînd în ambele urne facem egal numărul cazurilor favorabile (zece) însă în urna B mărim numărul cazurilor posibile cu 50, probabilitatea de a extrage o bilă albă din urna A va fi tot de $\frac{10}{100}$, adică de 10%, iar probabilitatea de a extrage o bilă albă din urna B va fi de $\frac{10}{150} = \frac{1}{15} = 0{,}066$ sau 6,6%.

Dacă în această din urmă ipoteză vom face serii de extracții din fiecare urnă — să zicem serii de cîte 100 extracții — punînd după fiecare extracție bila înapoi în urna din care s-a extras, vom obține valori *experimentale* de apariție a bilei considerate, care, în medie, vor oscila în jurul probabilităților teoretice pe care le-am determinat : pentru urna A numărul bilelor albe extrase în fiecare serie va oscila în jurul a 10%, iar pentru urna B, el va oscila în jurul a 6,6%.

Considerînd asemenea exemple, a fost formulată următoarea definiție a probabilității, corespunzătoare conceptului ei clasic :

Probabilitatea unui eveniment care se produce la întîmplare este raportul dintre numărul cazurilor favorabile și numărul cazurilor posibile, cu condiția ca toate cazurile să fie egal posibile.

Dacă într-o operație de masă evenimentul E se observă, în medie, de a ori într-un număr total de n cazuri (operații individuale), probabilitatea evenimentului ce se produce la întîmplare pe care o notăm cu $P(E)$, va fi egală cu raportul $\frac{a}{n}$, adică

$$P(E) = \frac{a}{n}. \qquad (I,2.1)$$

Aceasta este probabilitatea obținută pe cale experimentală pentru fenomenele de masă (a se vedea exemplul cu rebuturile de la p. 15). Mulți autori dau probabilităților astfel obținute, în mod impropriu, numele de „probabilitate statistică". Vom vedea mai departe sensul adevărat al acestei noțiuni.

Această definiție ne sugerează două observații importante care ne vor ajuta să înțelegem mai ușor teoria cîmpului de evenimente și a proprietăților lui, elaborată de Kolmogorov.

Prima observație. Numărul cazurilor favorabile este întotdeauna mai mic sau cel mult egal cu numărul cazurilor posibile; deci, probabilitatea dată de raportul (I,2.1) este întotdeauna *un număr pozitiv* cuprins între 0 și 1, inclusiv aceste valori. Cînd probabilitatea este egală cu 1 se zice că evenimentul se va întîmpla cu certitudine, îl putem considera ca un eveniment „sigur", deoarece aceasta înseamnă că numărul cazurilor favorabile este egal cu al celor posibile. De exemplu, într-o urnă cu 50 de bile albe, probabilitatea de a extrage o bilă albă este egală cu $\frac{50}{50} = 1$, căci numărul cazurilor favorabile este egal cu numărul cazurilor posibile.

Dacă nu avem cazuri favorabile, atunci probabilitatea are valoarea zero; deci, producerea evenimentului este imposibilă. De exemplu, dintr-o urnă cu 50 de bile negre, probabilitatea evenimentului de a fi extrasă o bilă albă este egală cu zero, evenimentul este imposibil.

Prin urmare: dacă valoarea probabilității este egală cu 1, $P(E) = 1$, evenimentul E se va întîmpla întotdeauna; dacă valoarea probabilității este egală cu 0, $P(E) = 0$, apariția evenimentului E poate fi considerată, practic, ca imposibilă.

Sînt cazuri cînd probabilitatea nulă reprezintă imposibilitatea realizării fenomenului. De exemplu, dacă știm că într-o urnă sînt 100 de bile albe și negre și că probabilitatea de a scoate o bilă albă este egală cu zero, înseamnă că în urnă nu este nici o bilă albă. În acest caz este imposibil să extragem o bilă albă.

Să presupunem acum că într-o urnă ideală sînt scrise pe bilete separate toate numerele reale. Ne întrebăm care este probabilitatea ca extrăgînd la întîmplare un bilet să obținem numărul 2 *). Dat fiind faptul că numărul numerelor reale este infinit, probabilitatea ca să obținem numărul 2 este egală cu zero. De data aceasta, cu toate că probabilitatea este nulă, nu există certitudinea că nu vom putea obține niciodată numărul 2.

Probabilitățile $P(E)$ verifică, deci, dubla inegalitate:

$$0 \leqslant P(E) \leqslant 1. \qquad (I,2.2)$$

Statisticienii obișnuiesc să considere un eveniment ca practic irealizabil, dacă probabilitatea ca el să se producă este mai mică decît un anumit număr (de exemplu 0,03). Determinarea acestui *prag de probabilitate*, după denumirea pe care o are în statistica matematică, variază de la caz la caz, după importanța și consecințele fenomenului considerat.

În domeniul producției industriale putem afirma uneori că o serie de produse fabricate este acceptabilă dacă probabilitatea de a obține un rebut este egală cu 1‰. Nu întotdeauna însă ne putem mulțumi cu o asemenea limită. Necesitățile practicii impun uneori un prag de probabili-

*) Problema poate fi formulată și în modul următor: pe axa numerelor reale se alege la întîmplare un număr; care este probabilitatea ca acest număr să fie 2?

tate egal cu $\frac{1}{100\,000}$ sau $\frac{1}{1\,000\,000}$ și chiar mai mic, echivalînd, practic, cu imposibilitatea apariției evenimentului, care, în general, nu este de dorit să apară în domeniul respectiv de activitate.

A doua observație. Dacă considerăm un eveniment E a cărui probabilitate de a se realiza este $P(E) = p$, nerealizarea acestui eveniment dă naștere unui alt eveniment E'. Ne propunem să calculăm probabilitatea de a se realiza evenimentul E', pe care o notăm cu $P(E') = q$. Pentru aceasta observăm că numărul cazurilor favorabile pentru evenimentul E' este egal cu cel al cazurilor nefavorabile pentru E, adică n-a, iar numărul cazurilor egal posibile este tot n, deoarece numărul total de cazuri a rămas același.

Prin urmare

$$P(E') = \frac{n-a}{n} = q. \qquad (I,2.3)$$

Se vede imediat, ținînd seama de egalitatea (I, 2.1), că avem relația simplă

$$p + q = \frac{a}{n} + \frac{n-a}{n} = \frac{n}{n} = 1. \qquad (I,2.4)$$

Evenimentele E și E' se numesc *evenimente contrare* sau *complementare*.

Dacă evenimentul E reprezintă extragerea unei bile albe din urna A, atunci evenimentul E' constituie extragerea unei bile de altă culoare decît cea albă (în cazul cînd urna conține bile de mai multe culori). Avem, deci, dacă într-o urnă sînt 40 de bile albe și 60 de bile negre :

$$a = 40, \qquad n = 100, \qquad n - a = 60,$$

$$p = \frac{40}{100}, \qquad q = \frac{60}{100},$$

$$p + q = \frac{40 + 60}{100} = 1,$$

de unde rezultă verificarea relației (I,2.4).

Definiția clasică a probabilității se bazează pe dubla ipoteză :

— Într-o operație de masă, rezultatul este constituit dintr-un număr bine determinat de n cazuri egal posibile, care *se exclud reciproc*, adică nu se pot întîmpla în același timp *într-o operație individuală* : o piesă nu poate fi în același timp și bună și rea, sau, extrăgînd o bilă dintr-o urnă cu bile de culoare albă și neagră, nu poate fi în același timp și albă și neagră etc.

— Evenimentul așteptat E se produce neapărat în operația de masă în a cazuri, numite favorabile lui E și nu se va produce în celelalte $n - a$ cazuri, numite nefavorabile lui E.

Ambelor ipoteze li se pot aduce obiecțiuni importante, a căror considerare a dus la o nouă formulare a noțiunii de probabilitate, mai cuprinză-

toare și mai precisă din punct de vedere științific, fiind legată și de viața reală, de practică.

Pentru a lămuri noțiunea de probabilitate în accepția ei modernă, trebuie să cunoaștem o altă noțiune folosită în calculul probabilităților și anume, noțiunea de *frecvență*.

Să presupunem că, aruncînd de 100 de ori o monedă, una dintre fețe a ieșit de 53 de ori. Raportul $\frac{53}{100}$ se numește *frecvență*. Așadar, frecvența înseamnă un *fapt experimental*. Dacă din n experiențe, un eveniment s-a verificat de m ori, raportul $\frac{m}{n}$ ne dă frecvența corespunzătoare apariției evenimentului considerat în seria de experiențe făcute. După cum se vede, *frecvența este și ea un număr pozitiv, cuprins între 0 și 1, inclusiv aceste valori*.

În primele trei exemple pe care le-am considerat în acest capitol, am pornit tocmai de la frecvențele unor evenimente (frecvența apariției pieselor conforme standardului, frecvența băieților născuți, frecvența bilelor albe în serii de extracții dintr-o urnă cu o anumită structură, cu condiția ca după fiecare extracție să se pună bila înapoi în urnă) într-o serie de mai multe experiențe și am constatat că ele oscilează în jurul unor anumite valori. Astfel, pentru fenomenele pentru care are loc această *stabilitate a frecvențelor*, am ajuns pe cale *experimentală, statistică*, la definirea noțiunii de probabilitate.

Prin urmare, *noțiunea teoretică de probabilitate a unui eveniment nu poate fi despărțită de noțiunea de frecvență. Calculul probabilităților se poate aplica numai fenomenelor pentru care există o stabilitate a frecvențelor în jurul probabilității*. Aceasta este baza legăturii teoriei matematice a probabilității cu lumea reală, cu practica. A proceda altfel înseamnă a rupe orice punte între calculele teoretice și statistica reală a evenimentelor și a ne posta, astfel, pe poziția de interpretare idealistă a rezultatelor unor experiențe *).

Reținem așadar constatarea că o definiție științifică a noțiunii de probabilitate trebuie să se refere în primul rînd la *comportarea reală* a fenomenului. Un fenomen trebuie să decurgă așa și nu altfel, datorită faptului că desfășurarea lui este condiționată de natura obiectivă a acestui fenomen, de întregul complex de cauze interdependente care-i determină legea obiectivă de dezvoltare și nu fiindcă „*noi ne așteptăm* tocmai la această desfășurare a lui, fiindcă *noi am fi mirați*, dacă el s-ar desfășura altfel”.... **).

„Posibilitatea nu este deci expresia gradului subiectiv de încredere a omului în producerea evenimentului, ci caracterizarea legăturii obiectiv existente între condiții și eveniment” ***).

---

*) Vezi în acest sens A. I. Hincin, *Metoda funcțiilor arbitrare și lupta împotriva idealismului în teoria probabilităților*. Analele Romîno-Sovietice, matematică-fizică, 2 (1953).

**) A. I. Hincin, *Op. cit.*

***) Vezi în acest sens *Bazele filozofiei marxiste*. Ed. politică, București, 1959, p. 256.

Multe din fenomenele naturii și ale vieții sociale se desfășoară după *o lege de probabilitate*; ca orice lege a științei și legea de probabilitate are un caracter *obiectiv*, *necesar* și *general*.

Descoperirea legilor de probabilitate este rezultatul unui proces îndelungat al activității de abstractizare a gîndirii științifice în domeniul calculului probabilităților.

Faptul că o lege de probabilitate este obiectivă, înseamnă că fenomenul studiat se va supune acestei legi independent de voința noastră; faptul că este generală, înseamnă că ea este proprie tuturor fenomenelor și proceselor de anumit gen. Aceste caractere ale legilor de probabilitate o apropie de celelalte legi ale diferitelor științe ale naturii și societății.

În ceea ce privește caracterul de *necesitate* al legilor de probabilitate, el capătă un anumit sens în teoria probabilităților. Se știe că legile fenomenelor naturii și societății, formulate de diferite științe, sînt necesare în sensul că la fiecare realizare a unui ansamblu de condiții A are loc întotdeauna evenimentul B; *legăturile legice dintre fenomene acționează*, deci, *în anumite condiții, în mod obligatoriu, cu forța necesității*. De exemplu, legea căderii corpurilor descoperită de Galileu acționează cu forța necesității: toate corpurile și întotdeauna cad pe pămînt cu o accelerație de 9,8 m/s.

Există însă o vastă categorie de fenomene pentru care la îndeplinirea unui ansamblu de condiții A *nu rezultă cu necesitate evenimentul* B, însă, *dacă avem o repetare a unui număr mare de experiențe, cu îndeplinirea ansamblului de condiții* A, *se manifestă o anumită legitate exprimată prin oscilarea frecvenței în jurul unui număr care reprezintă elementul stabil, o valoare caracteristică obiectivă a fenomenului studiat, pe care o notăm prin* p.

Vom spune că, pentru o asemenea categorie de fenomene, *probabilitatea ca la îndeplinirea sistemului de condiții* A *să se realizeze evenimentul* B *este egală cu* p.

În grupuri formate de un număr mare de probe, frecvențele reprezintă *valorile experimentale, statistice, ale probabilității* p *care este o constantă obiectivă*, depinzînd de natura fenomenului considerat și de sistemul de condiții A.

După cum se vede, probabilitatea unui eveniment are sens atît timp cît *condițiile* în care are loc operația de masă respectivă rămîn neschimbate; orice schimbare a acestor condiții atrage după sine schimbarea probabilității, schimbarea legii de probabilitate a fenomenului. Așadar, acțiunea legilor de probabilitate depinde de anumite condiții, formele lor de manifestare schimbîndu-se în diferite condiții concrete.

Precizările de mai sus ne permit să tragem concluzia că o lege obiectivă de probabilitate se manifestă concret ca o *legitate statistică* a ansamblului de fenomene pentru a cărei descoperire trebuie examinate un număr mare de cazuri individuale întîmplătoare. De asemenea, ele ne permit să scoatem în evidență o primă insuficiență a definiției clasice a probabilității, aceea de a fi ruptă de lumea reală.

În ceea ce privește o altă importantă obiecție care se aduce definiției clasice, nu este nici ea mai puțin categorică.

Scanned with OKEN Scanner

După cum știm, definiția clasică a probabilității, ca fiind un raport dintre numărul cazurilor favorabile și numărul cazurilor posibile, se bazează esențial pe discriminarea cazurilor egal posibile. Dacă vom considera, de exemplu, un zar perfect construit din punct de vedere geometric și perfect omogen din punct de vedere fizic, apariția fiecărei fețe a zarului constituie un caz egal posibil. Definiția este deci aplicabilă. Să presupunem însă (ceea ce este conform cu realitatea) că zarul nu este perfect din mai multe puncte de vedere. În acest caz este evident că dispare orice posibilitate de clasificare a feței zarului în cazuri egal posibile. Prin urmare, nici chiar atunci cînd avem un număr finit de cazuri posibile, discriminarea lor în cazuri *egal* posibile*) nu se poate face și aceasta, de cele mai multe ori.

Insuficiența definiției clasice a probabilității se vădește și mai mult cînd este vorba de un număr infinit de cazuri. Se pot da nenumărate exemple în care probabilitatea apariției unui eveniment — în sensul în care este precizată în definiția clasică — nu poate fi determinată, nu numai din cauza imposibilității de a discrimina numărul cazurilor egal posibile, ci și din cauza dificultății de a determina, în general, numărul cazurilor. Acest lucru se întîmplă, de exemplu, în cazul cînd am vrea să determinăm probabilitatea de apariție a unui număr prim, luînd la întîmplare un număr din șirul numerelor naturale, sau, probabilitatea ca o curbă închisă, de lungime dată, să închidă o arie mai mică decît un număr dat etc.

Se mai poate invoca încă un motiv pentru care definiția clasică a probabilității nu poate fi admisă și anume că ea nu poate fi aplicată unei largi categorii de fenomene naturale și sociale în studiul cărora apare imposibilitatea de a fi determinat numărul cazurilor favorabile, respectiv, nefavorabile. De exemplu, fără a recurge la frecvență nu se poate determina *apriori* probabilitatea ca dintr-un număr de N persoane de aceeași vîrstă peste un an să fie n în viață, sau, probabilitatea ca peste o anumită regiune să cadă grindină într-un interval de timp dat etc. Or, dacă calculul probabilităților nu s-ar putea aplica în studiul unor fenomene și procese care au loc în natură și societate, el ar înceta de a mai fi o știință, limitîndu-se doar a fi un studiu matematic, lipsit de aplicații, al jocurilor de noroc.

A releva lipsurile definiției clasice a probabilității nu înseamnă însă a aluneca pe panta unei poziții agnostice, renunțînd, în fața dificultăților teoretice și practice, la posibilitatea determinării probabilității pe baza legilor obiective de dezvoltare a fenomenelor, așa cum s-a întîmplat cu matematicianul Mises **).

Mises a definit probabilitatea ca limită a frecvenței, atunci cînd fenomenul ar fi supus unui număr nelimitat de probe. Potrivit teoriei frecven-

---

*) Procedeul este aplicabil numai pentru scheme probabilistice construite cu mare atenție. De exemplu, o urnă în care sînt așezate bile identice ca formă și greutate, pentru care se folosește un sistem de extracție astfel conceput încît bilele să nu poată fi deosebite între ele; o ruletă etc.

**) R. Mises (1883), matematician german. A fost profesor la universitatea din Berlin în perioada 1920—1933. Cele mai importante lucrări ale lui Mises sînt din domeniul calculului probabilităților.

țelor concepută de el, dacă notăm prin f_n frecvența corespunzătoare pentru n probe, avem

$$P = \lim_{n\to\infty} f_n. \qquad (I,2.5)$$

Împotriva acestei definiții se pot aduce argumente hotărîtoare pe care însă nu este cazul să le expunem aici. Progresul realizat în ultimul timp în teoria probabilităților a eliminat definitiv teoria agnostică a lui Mises.

3. Schema lui Kolmogorov *)

Obiectele de bază folosite în axiomatizarea teoriei probabilităților făcută de Kolmogorov sînt *evenimentele* și *probabilitățile* respective.

Legătura pe care am făcut-o înainte între probabilitate și frecvență a dus la constatarea că evenimentele corespunzătoare diferitelor experiențe se bucură de unele proprietăți care pot fi formulate matematic, formînd împreună o *structură* pe care Kolmogorov a denumit-o *cîmp de evenimente*.

PRIMA AXIOMĂ. *Unei experiențe îi corespunde întotdeauna un cîmp de evenimente.*

Aceasta este prima axiomă din schema Kolmogorov.

Să luăm experiența clasică a aruncării zarului. Referitor la această experiență pot fi considerate mai multe evenimente; putem preciza chiar că ele sînt în număr de $64 = 2^6$.

Într-adevăr, avem în primul rînd un *eveniment sigur*. Atunci cînd aruncăm un zar sîntem siguri că va ieși una din cele șase fețe ale zarului. Acest eveniment îl putem considera evenimentul $(1, 2, 3, 4, 5, 6)$.

În mod analog putem considera și alte cinci grupe de evenimente de categorii diferite, constînd, succesiv:

I. apariția uneia din față (avem șase evenimente)

$$(1),\ (2),\ \ldots,\ (6);$$

II. apariția uneia din două fețe:

$$(1,\ 2),\ (1,\ 3),\ \ldots,\ (5,\ 6);$$

ele sînt în număr de $C_6^2 = 15$;

III. apariția uneia din trei fețe, deci, evenimente de forma $(1, 2, 3)$. Totalitatea lor este

$$(1,\ 2,\ 3),\ (1,\ 2,\ 4),\ \ldots,\ (4,\ 5,\ 6)$$

deci, în număr egal cu $C_6^3 = 20$;

*) A. N. Kolmogorov (n. 1903), matematician sovietic, membru al Academiei de Științe din U.R.S.S., a adus contribuții importante în multe ramuri ale matematicii, ca teoria funcțiilor de variabilă reală, topologie, algebră, logică matematică etc. Cele mai însemnate lucrări ale lui Kolmogorov sînt însă în domeniul calculului probabilităților, unde este autorul axiomatizării acestei discipline matematice, și în statistica matematică, unde este unul din fondatorii statisticii neparametrice. De asemenea, a adus o contribuție fundamentală în teoria proceselor stohastice.

IV. apariția uneia din patru fețe, evenimente de forma (1, 2, 3, 4), al căror număr este de $C_6^4 = 15$ și în sfîrșit,

V. apariția uneia din cinci fețe, evenimente de forma (1, 2, 3, 4, 5) sau (1, 2, 3, 4, 6) etc., al căror număr este de $C_6^5 = 6$.

Avem, așadar, $1 + 6 + 15 + 20 + 15 + 6 = 63$ evenimente la care mai adăugăm și *evenimentul imposibil* care constă, evident, în faptul imposibil ca să nu iasă nici una din fețele zarului. Obținem astfel cele 64 evenimente pe care le putem considera în cîmpul de evenimente, corespunzător experienței de aruncare o singură dată a zarului.

Examinarea acestui cîmp particular de evenimente ne conduce la unele proprietăți generale pe care le întîlnim în oricare alte exemple de evenimente.

Într-un cîmp de evenimente putem distinge următoarele categorii de evenimente (ele rezultă și din exemplul cu zarul) care conturează proprietățile generale ale unui cîmp de evenimente:

a) *Evenimente echivalente.* Dacă din cîmpul de evenimente ale experienței cu zarul considerăm evenimentele (1) și (1, 2), observăm că în cazul cînd are loc primul eveniment, are loc și evenimentul al doilea. Deci evenimentul (1) implică (1, 2).

Implicația a două evenimente A și B din același cîmp de evenimente se notează cu semnul $\sqsubset$.

Vom putea scrie:

$$A \sqsubset B.$$

Dacă realizarea evenimentului A implică realizarea evenimentului B și invers

$$B \sqsubset A,$$

vom spune că evenimentele A și B sînt *echivalente.*

b) *Evenimente reunite.* Operația de *reuniune* se notează cu semnul $\cup$. În general, pentru orice cîmp de evenimente dacă considerăm două evenimente oarecare A și B, $A \cup B$, adică sau A sau B, reprezintă de asemenea un eveniment al cîmpului.

De exemplu,

$$(1) \cup (2) = (1, 2).$$

c) *Evenimente intersectate.* Intersecția se notează cu semnul $\cap$. În general, dacă A și B sînt două evenimente din același cîmp vom nota cu

$$A \cap B$$

evenimentul care constă în realizarea simultană a evenimentului A și a evenimentului B.

Evenimentul $A \cap B$ îl vom citi A și B sau A *intersectat* cu B.

De exemplu, să considerăm evenimentele (1, 2, 3), (1,3); evenimentul care constă în realizarea simultană și a evenimentului (1, 2, 3) și a evenimentului (1, 3) este evident, evenimentul (1, 3), și-l vom nota

$$(1, 2, 3) \cap (1,3) = (1, 3).$$

Dacă evenimentele intersectate se exclud reciproc — asemenea evenimente se numesc *incompatibile* — operația de intersecție conduce tot la un eveniment din cîmpul considerat și anume la un eveniment *imposibil*, pe care-l vom nota cu Φ.

De exemplu,

$$(1, 2) \cap (5, 6) = \Phi.$$

d) *Evenimente complementare.* Dacă două evenimente *incompatibile* din același cîmp de evenimente A și B, reunite, ne duc tot la un eveniment al aceluiași cîmp, *eveniment sigur* (E), se spune că aceste evenimente sînt *complementare*.

Deci

$$A \cap B = \Phi \quad \text{și} \quad A \cup B = E.$$

Notăm aceasta

$$A = \complement B,$$

unde semnul $\complement$ arată că evenimentele sînt complementare.

e) *Evenimente elementare.* Evenimentele (1), (2), (3), (4), (5), (6) care după notația noastră constă în ieșirea fețelor 1, 2, ..., 6 ale zarului, sînt *evenimente elementare*.

Un eveniment A al unui cîmp este *elementar* dacă, față de un alt eveniment oarecare B al cîmpului, are una din situațiile:

1°. A este *incompatibil* cu B, de exemplu apariția feței 1 a zarului exclude apariția oricărei alte fețe; sîntem în cazul evenimentelor *incompatibile*.

2°. A implică pe B, adică

$$A \subset B.$$

Este evident că reuniunea evenimentelor elementare ne dă evenimentul sigur.

A DOUA AXIOMĂ. *Fiecărui eveniment A al cîmpului de evenimente îi corespunde un număr real, nenegativ, $P(A)$ care se numește probabilitate.*

După cum am văzut din exemplele considerate la începutul acestui capitol, experiența ne arată că frecvența este apropiată de probabilitate. Este natural, deci, să admitem că probabilitatea este un număr *real*, *nenegativ*, deoarece avem întotdeauna dubla inegalitate

$$0 \leqslant \frac{m}{n} \leqslant 1,$$

în care m reprezintă un număr care arată de cîte ori se verifică evenimentul A, cînd avem o repetare a unui număr mare de experiențe (n), cu îndeplinirea ansamblului de condiții identice A.

A TREIA AXIOMĂ. *Probabilitatea evenimentului sigur este egală cu* 1.

Deci

$$P(E) = 1,$$

în care evenimentul sigur este notat cu E.

A PATRA AXIOMĂ. *Probabilitatea evenimentului reunit este egală cu suma probabilităților fiecărui eveniment în parte, cu condiția ca evenimentele componente să fie incompatibile între ele.*

Avem

$$P(A \cup B) = P(A) + P(B).$$

Adunarea probabilităților pentru evenimentele incompatibile este un fapt verificat experimental de multă vreme. În unele tratate de calculul probabilităților, apărute înainte de axiomatizarea acestei discipline matematice, ea apare sub denumirea de „teoremă a probabilităților totale".

Fie, de exemplu, $q(x, x+1)$ probabilitatea ca o persoană de vîrstă x să moară la o vîrstă cuprinsă între x și $x+1$ ani; $q(x+1, x+2)$ probabilitatea ca persoana să moară între $x+1$ și $x+2$ ani, iar $q(x, x+2)$, probabilitatea ca aceeași persoană să moară între x și $x+2$ ani; avem

$$q(x, x+2) = q(x, x+1) + {}_1q(x+1, x+2).$$

Această relație și-a găsit aplicații în asigurări de viață. De altfel această relație dintre probabilități este strîns legată de proprietatea corespunzătoare pentru frecvențe. Cum pentru frecvență această egalitate este întotdeauna îndeplinită, este clar că dacă nu am admite și pentru probabilități egalitatea corespunzătoare, nu ar putea să mai subziste legătura confirmată de experiență, între probabilitate și frecvență.

*

Să revenim acum la definiția clasică a noțiunii de probabilitate, reexaminînd-o în lumina axiomatizării făcută de Kolmogorov. Ne putem ușor convinge că alegerea unei probabilități $P(A)$ egală cu raportul dintre numărul cazurilor favorabile și numărul cazurilor posibile, cînd avem un număr finit de cazuri, nu vine în contradicție cu cele patru axiome ale lui Kolmogorov, cu *condiția* însă ca, în fiecare caz concret pentru care se cere determinarea probabilității, să putem face *discriminarea cazurilor egal posibile.*

Vom expune în continuare problemele legate de determinarea numerică a probabilităților, folosind, în genere, schemele clasice, corectate prin axiomatizarea elaborată de Kolmogorov.

Axiomatizarea lui O. Onicescu. Fenomene din ce în ce mai numeroase impun astăzi o lărgire a cadrului axiomatic al teoriei probabilităților. Încă din 1954, A. N. Kolmogorov a semnalat dificultățile pe care le întîlnește teoria lui și necesitatea de a o lărgi punîndu-i la bază nu o mulțime de evenimente elementare reprezentate prin puncte ale unui spațiu, în general convențional, ci o mulțime de evenimente care constituie o latice, cu element unitate (evenimentul total): complementată și numărabil aditivă.

Evenimentele elementare ale cîmpului în axiomatica lui Kolmogorov au în general, cum a observat el însuși, probabilitatea nulă.

Evenimentele laticei, evenimentele propriu-zise ale cîmpului, au probabilități în genere pozitive și aditive.

Considerarea acestor cîmpuri de evenimente reprezentate, ca structuri matematice, prin algebre Boole susceptibile de o măsură normată, nu rezolvă încă din punctul de vedere al prof.

Onicescu, problema constituirii teoriei probabilității. Este necesară reprezentarea mărimilor ce corespund evenimentelor cîmpului. Aceste mărimi, dacă sînt numerice, sînt funcții σ-aditive de evenimente, nule pentru evenimentele cu probabilitatea nulă. Funcțiile cu astfel de proprietăți sînt *funcțiile sumă* ale lui Onicescu:

$$F(\omega) = 0 \text{ dacă } P(\omega) = 0,$$
$$F(\omega_1 \cup \omega_2 \cup \ldots) = \sum_j F(\omega_j),$$

dacă $\omega_1 \cap \omega_2 = 0$, $\omega_1 \cap \omega_3 = 0$ etc.
Funcțiile sumă elementare au forma

$$F(\omega) = \sum_{j=1}^{k} \frac{F(\omega_j)}{P(\omega_j)} P(\omega \cap \omega_j), \qquad (I,3.1)$$

unde $\omega_1, \omega_2, \ldots, \omega_k$ constituie un sistem *complet* de evenimente *disjuncte* două cîte două

$$\omega_j \cap \omega_h = 0,$$
$$\omega_1 \cup \omega_2 \cup \ldots \cup \omega_k = E.$$

Orice funcție sumă este limită uniformă de funcții simple. Dacă ne găsim în cadrul mai restrîns al teoriei lui Kolmogorov, funcțiile sumă coincid cu integralele

$$\int f(\xi)\, P(d\xi),$$

unde $f(\xi)$ este o variabilă alcatoare oarecare ω integrabilă.

Proprietatea fundamentală a funcțiilor sumă este următoarea:

$F(\omega)$ fiind o funcție sumă, există un eveniment θ pentru care

$$F(\omega) \geqslant 0 \text{ cînd } \omega < \theta$$
$$F(\omega) < 0 \text{ cînd } \omega < \theta',$$

θ' fiind evenimentul complementar lui θ.

Principalele teoreme ale teoriei probabilităților se pot enunța și demonstra cu o deosebită ușurință folosind funcțiile sumă. Ele au proprietăți deosebit de simple și de comode în calcul.

Cum orice funcție sumă este limita uniformă a unui șir de funcții sumă simple de forma (I,3.1), este suficient, de cele mai multe ori, să demonstrăm valabilitatea unei teoreme pentru funcțiile sumă de forma (I, 3.1), pentru a o deduce apoi, în mod simplu, pentru o funcție sumă oarecare.

Teoria proceselor aleatoare se prezintă cu ajutorul funcțiilor sumă mai simplu decît cu ajutorul funcțiilor și variabilelor aleatoare obișnuite.

I. Cuculescu a dezvoltat teoria martingalei cu ajutorul funcțiilor sumă obținînd o prezentare deosebit de simplă și sugestivă în aceiași vreme a acestei teorii.

Gh. Marinescu a studiat spațiul Banach al funcțiilor sumă ale unui cîmp.

Scanned with OKEN Scanner

Capitolul II

Determinarea probabilităţilor. Reguli de calcul pentru evenimente în număr finit

1. Consideraţii generale

Sistemul de axiome elaborat de Kolmogorov nu determină numeric probabilităţile evenimentelor. A. N. Kolmogorov arată *) că probabilitatea $P(A)$ corespunzătoare evenimentului A, în afară de îndeplinirea axiomelor teoriei probabilităţilor, trebuie să se bucure şi de următoarele proprietăţi :

1°. Dacă fenomenul se desfăşoară de n ori în condiţii identice date de sistemul de condiţii A şi evenimentul A se verifică de m ori, trebuie să fim siguri că, practic, frecvenţa $\frac{m}{n}$ diferă foarte puţin de $P(A)$, dacă n e suficient de mare.

2°. Dacă $P(A)$ este o valoare foarte mică, să fim siguri că într-o singură experienţă evenimentul A nu se verifică.

În anumite cazuri, determinarea probabilităţii se poate face direct. A. I. Hincin **) a conceput o metodă pe care a numit-o *metoda funcţiilor arbitrare*, care face posibilă deducerea stabilităţii frecvenţelor unor anumite fenomene mecanice şi, în consecinţă, determinarea probabilităţii respective, din studiul legilor de dezvoltare a acestor fenomene.

---

*) Vezi A. N. Kolmogorov, *Grundbegriffe der Wahrscheinlichkeitsrechnung*, Berlin, 1933.

**) Vezi A. I. Hincin, *Metoda funcţiilor arbitrare şi lupta împotriva idealismului în teoria probabilităţilor*. Analele Romîno-Sovietice, matematică-fizică 2 (1953). A. I. Hincin a lucrat în domeniul analizei matematice, teoriei numerelor şi calculului probabilităţilor. În teoria probabilităţilor el a adus contribuţii deosebit de importante în studiul legilor indefinit divizibile, în teorema logaritmului iterat, ca şi în studiul proceselor staţionare etc. Metodele şi rezultatele teoriei probabilităţilor au fost folosite pe larg de Hincin în fizica statistică şi teoria informaţiei.

Metoda funcțiilor arbitrare a lui Hincin. Vom expune metoda funcțiilor arbitrare a lui Hincin pe exemplul simplu al ruletei. Metoda este însă aplicabilă la numeroase fenomene fizice.

Bila, considerată ca un punct material, se mișcă pe un cerc, a cărui lungime o considerăm egală cu 1. Mișcarea începe de la un punct bine determinat și are loc în același sens. Vom neglija frecarea și ceilalți factori care micșorează viteza bilei, presupusă constantă în tot timpul mișcării. Cercul se împarte în $2n$ sectoare egale, vopsite alternativ în roșu și negru. Vom nota prin A evenimentul constînd în oprirea bilei pe o porțiune roșie și prin B evenimentul contrar (oprirea pe o porțiune neagră). După trecerea unui timp t bila se oprește brusc. Se pune problema demonstrării stabilității fenomenului A într-o lungă serie de experiențe.

Să însemnăm prin v viteza inițială a bilei și să presupunem că frecvența valorilor vitezei inițiale cuprinse între v_1 și v_2 este

$$\int_{v_1}^{v_2} f(v)\,dv,$$

integrala fiind luată în sensul lui Lebesgue, iar funcția $f(v)$ îndeplinind condițiile

$$f(v) \geqslant 0, \qquad (0 \leqslant v < +\infty), \qquad \int_0^{\infty} f(v)\,dv = 1.$$

Spațiul parcurs de bilă în timpul t este vt. El cuprinde un număr întreg de arce de lungime $\frac{1}{n}$ și o anumită parte suplementară λ. Avem deci

$$vt = \frac{k}{n} + \lambda, \qquad \left(0 \leqslant \lambda < \frac{1}{n}\right).$$

Evenimentul A se produce dacă $\lambda < \frac{1}{2n}$ și evenimentul B dacă $\lambda \geqslant \frac{1}{2n}$. Rezultă că evenimentul A se produce întotdeauna cînd viteza inițială aparține unuia din segmentele

$$\left[\frac{k}{nt}, \frac{k+\frac{1}{2}}{nt}\right], \qquad (k = 0, 1, 2, \ldots).$$

Însemnînd prin $P(A)$ frecvența vitezelor corespunzătoare cazurilor cînd evenimentul A se produce, avem

$$P(A) = \sum_{k=0}^{\infty} \int_{\frac{k}{nt}}^{\frac{k+\frac{1}{2}}{nt}} f(v)\,dv.$$

În mod analog, avem pentru evenimentul B

$$P(B) = \sum_{k=0}^{\infty} \int_{\frac{k+\frac{1}{2}}{nt}}^{\frac{k+1}{nt}} f(v)\,dv.$$

Prin urmare

$$|P(A) - P(B)| = \left| \sum_{k=0}^{\infty} \int_{\frac{k}{nt}}^{\frac{k+\frac{1}{2}}{nt}} f(v)\,dv - \sum_{k=0}^{\infty} \int_{\frac{k+\frac{1}{2}}{nt}}^{\frac{k+1}{nt}} f(v)\,dv \right| \leqslant$$

$$\leqslant \sum_{k=0}^{\infty} \int_{\frac{k}{nt}}^{\frac{k+\frac{1}{2}}{nt}} \left| f\left(v + \frac{1}{2nt}\right) - f(v) \right| dv = \int_0^{\infty} \left| f\left(v + \frac{1}{2nt}\right) - f(v) \right| dv .$$

Rezultă

$$P(A) \to P(B), \text{ pentru } t \to \infty .$$

Cum

$$P(A) + P(B) = 1,$$

ne convingem că, oricare ar fi legea de repartiție a vitezelor inițiale, frecvența apariției evenimentului A va avea o valoare apropiată de 1/2.

Este cazul să arătăm că în literatura noastră de specialitate s-au dat exemple de stabilitate a frecvențelor înaintea lui A. I. Hincin *). Să considerăm astfel un sistem de m urne U_i $(\lambda_i\, \mu_i)$, $(i = 1,2,\ldots,m)$, conținînd fiecare λ_i bile albe și μ_i bile negre. Presupunem că avem $\lambda_1 = 0$, $\mu_m = 0$, $\lambda_i > 0$, $\mu_i > 0$, $(i \neq 1, i \neq m)$. Din aceste urne efectuăm extracții succesive după următoarea regulă: dacă bila extrasă din urna U_i este albă, extracția următoare este făcută din urna U_{i-1} iar dacă a fost neagră, extracția următoare este făcută din U_{i+1}.

Dacă însemnăm prin α numărul de bile albe obținut în n extracții succesive făcute după această regulă, se demonstrează că

$$\lim_{n\to\infty} \frac{\alpha}{n} = \frac{1}{2}.$$

Demonstrația este extrem de simplă. Fie α numărul bilelor albe și β numărul bilelor negre obținute în n extracții

$$\alpha + \beta = n.$$

Prima extracție fiind efectuată din urna U_k, presupunem că după n extracții am ajuns la urna U_j.

Dacă $j > k$, am obținut $j-k$ bile negre în plus față de numărul bilelor albe:

$$\beta - \alpha = j - k.$$

Rezultă deci

$$\alpha = \frac{n}{2} - \frac{j-k}{2}, \qquad \beta = \frac{n}{2} + \frac{j-k}{2},$$

$$\left| \frac{\alpha}{n} - \frac{1}{2} \right| < \frac{|j-k|}{2n}.$$

*) O. Onicescu și G. Mihoc, *Sopra le leggi-limite delle probabilità*, Giornale dell'Istituto Italiano degli Attuari, Roma, *1* (1936).

O. Onicescu și G. Mihoc, *Asupra unui șir de urne dispuse circular*. Analele științifice ale Universității din Iași, XXIV (1938).

Scanned with OKEN Scanner

Deoarece

$$|j-k| < m$$

obținem

$$\left|\frac{\alpha}{n} - \frac{1}{2}\right| < \frac{m}{2n}.$$

Deci

$$\lim_{n\to\infty} \frac{\alpha}{n} = \frac{1}{2}$$

Prin urmare, frecvența $\frac{\alpha}{n}$ va avea o valoare apropiată de $\frac{1}{2}$ pentru n foarte mare, tinzînd chiar către $\frac{1}{2}$ pentru $n\to\infty$.

Să admitem că într-o urnă sînt bile de culorile A_1, A_2,...,A_k și că facem extracții din urnă după următoarea regulă: dacă bila extrasă este de culoarea A_i, punem în urnă o bilă de culoarea A_{i+1}, $(i = 1, 2,..., k-1)$, iar dacă bila extrasă este de culoarea A_k punem în urnă o bilă de culoarea A_1. Efectuăm n extracții succesive după această regulă. Se demonstrează la fel ca mai sus, dacă însemnăm prin α_i numărul de bile de culoarea A_i scoase în n extracții, că avem

$$\lim_{n\to\infty} \frac{\alpha_i}{n} = \frac{1}{k}, \qquad (i = 1,2,...,k).$$

Urmînd aceeași cale, se pot construi exemple de scheme probabilistice pentru care frecvențele tind în mod riguros, cînd numărul experiențelor crește necontenit, către orice număr dat mai înainte, cuprins între 0 și 1.

În majoritatea cazurilor însă, pornind de la analiza cauzelor care determină fenomenul, vom ajunge la determinarea probabilității P folosind faptul că la repetarea experiențelor în sistemul de condiții A, frecvențele se grupează în jurul lui P. Vom proceda astfel, bineînțeles, dacă analiza prealabilă a cauzelor fenomenului n-a condus la o anumită valoare a probabilității sau la o anumită formă a probabilității ca funcție, de exemplu, de un parametru numeric care urmează să fie determinat.

Să presupunem astfel că admitem probabilitatea de a ieși fața 1 a unui zar egală cu $\frac{1}{6}$ și, făcînd numeroase serii de aruncări, constatăm că frecvențele se depărtează sistematic de $\frac{1}{6}$. Știind că obținerea acestui rezultat are cauze obiective perfect determinate, putem afirma că probabilitatea evenimentului nu este egală cu $\frac{1}{6}$; în acest caz valoarea probabilității *va fi determinată empiric, în raport cu frecvențele obținute.*

2. Reguli de calcul

Ne vom opri acum asupra cîtorva exemple clasice de determinare numerică a probabilităților, în care stabilirea numărului cazurilor egal posibile nu prezintă dificultăți.

EXEMPLUL 1. *Se aruncă o monedă în sus. Care este probabilitatea ca moneda să cadă pe o față aleasă mai înainte?*

Observăm că moneda are două fețe, deci avem două cazuri posibile. Se pune însă întrebarea dacă cele două fețe sînt *egal* posibile. Răspunsul este afirmativ pentru că simetria pe care o observăm la forma în alcătuirea monedei ne face să considerăm că nu există din punctul nostru de vedere o față privilegiată față de cealaltă.

Numărul cazurilor favorabile este, evident, egal cu 1, adică fața pe care am ales-o la început.

Probabilitatea căutată este prin urmare $\frac{1}{2}$.

EXEMPLUL 2. *Se aruncă un zar. Care este probabilitatea ca să iasă fața 6 a zarului?*

Zarul are șase fețe. Avem prin urmare șase cazuri posibile. Simetria pe care o observăm în construcția zarului ne îndreptățește să afirmăm că aceste cazuri sînt și egal posibile.

Întrucît avem o singură față 6, este evident că numărul cazurilor favorabile este egal cu 1.

Probabilitatea căutată este, deci, $\frac{1}{6}$.

EXEMPLUL 3. *Care este probabilitatea de a ieși cel puțin odată marca unei monede aruncînd-o de două ori?*

Numărul cazurilor egal posibile se obține simplu: din prima aruncare a monedei poate ieși marca sau banul, din a doua la fel. Diferitele posibilități pe care le putem obține sînt:

Prima aruncare:	A doua aruncare:
marca	marca
marca	banul
banul	marca
banul	banul

Avem deci patru cazuri egal posibile, din care numai primele trei sînt favorabile, deoarece marca a ieșit cel puțin odată în fiecare.

Probabilitatea căutată este:

$$P = \frac{3}{4}.$$

EXEMPLUL 4. *Care este probabilitatea P ca aruncînd de trei ori un zar să obținem cel puțin odată fața 6?*

Vom determina mai întîi numărul cazurilor egal posibile. Observăm că la prima aruncare sînt posibile șase cazuri, deoarece poate ieși oricare din fețele 1, 2, 3, 4, 5, 6. La a doua sînt $6 \times 6 = 36 = 6^2$ cazuri posibile, deoarece fiecare din cele șase cazuri referitoare la prima aruncare se poate asocia cu oricare din cele șase ce rezultă după aruncarea a doua. În sfîrșit, după cea de a treia aruncare avem $36 \times 6 = 6^3 = 216$ cazuri posibile.

Pentru determinarea numărului de cazuri favorabile se poate forma un tablou analog cu cel din problema precedentă. Acest procedeu este însă greoi și de aceea se poate utiliza probabilitatea *complementară* din

cîmpul de evenimente al experienței noastre și anume, probabilitatea ca din trei aruncări să iasă de fiecare dată una din fețele 1, 2, 3, 4, 5. Numărul cazurilor favorabile este, în acest caz, $5 \times 5 \times 5 = 5^3$. Numărul cazurilor egal posibile fiind 6^3, rezultă că probabilitatea evenimentului complementar (a evenimentului ca să iasă numai fețele 1, 2, 3, 4, 5), este $\frac{5^3}{6^3}$.

Or, după cum știm, reuniunea evenimentelor incompatibile complementare dintr-un cîmp de evenimente ne dă un alt eveniment din același cîmp de evenimente — evenimentul sigur — ceea ce înseamnă, în exemplul nostru, că

$$P + \frac{5^3}{6^3} = 1 .$$

De aici rezultă că probabilitatea căutată P ca aruncînd de trei ori zarul să obținem cel puțin odată fața 6 este:

$$P = 1 - \frac{5^3}{6^3} = \frac{91}{216},$$

Acest exemplu ne arată că determinarea probabilității prin calculul direct al numărului de cazuri favorabile și posibile prezintă uneori dificultăți și devine cu atît mai anevoiasă cu cît numărul cazurilor este mai mare, iar discriminarea lor mai greu de făcut.

EXEMPLUL 5. *Se aruncă un zar de două ori. Care este probabilitatea ca suma punctelor obținute din prima și a doua aruncare să fie un număr dat dinainte?*

Este evident că suma punctelor obținute nu poate fi decît unul din numerele 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12. Să notăm probabilitățile corespunzătoare fiecărei sume prin:

$$P_2, P_3, P_4, \ldots, P_{11}, P_{12} .$$

Probabilitatea P_2 se determină observînd că suma $2 = 1 + 1$, se obține dacă la prima aruncare iese fața 1 iar la a doua, la fel fața 1. Avem un singur caz favorabil pentru suma egală cu numărul 2, iar numărul cazurilor posibile este 36, deoarece la fiecare față a zarului din prima aruncare se pot asocia cele șase fețe ale zarului din a doua aruncare. Probabilitatea căutată este deci:

$$P_2 = \frac{1}{36} .$$

Pentru P_3 avem următoarele cazuri favorabile:

a) cînd la prima aruncare iese fața 1, iar la a doua fața 2;

b) cînd la prima aruncare iese fața 2, iar la a doua fața 1.

Numărul cazurilor favorabile este deci 2; prin urmare, valoarea probabilității P_3 este

$$P_3 = \frac{2}{36} .$$

În mod analog găsim valorile probabilităților P_4, P_5, P_6, ..., P_{11}, P_{12}.

Vom avea următorul tablou final al valorilor care le iau sumele și al probabilităților corespunzătoare:

Valorile numerelor	2	3	4	5	6	7	8	9	10	11	12
Probabilitățile	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

Observăm că cea mai mare probabilitate de a se realiza o are suma punctelor egală cu șapte.

EXEMPLUL 6. *O urnă conține 80 de bile albe și 20 de bile negre. Din această urnă se fac patru extracții succesive, după fiecare extracție punîndu-se bila înapoi în urnă. Care este probabilitatea ca din cele patru bile astfel extrase, două să fie albe?*

Pentru a rezolva această problemă în sensul definiției clasice a probabilității, vom raționa după cum urmează:

Numărul cazurilor egal posibile în cele patru extracții este egal cu $100^4 = 100\,000\,000$.

Într-adevăr, la prima extracție putem obține oricare din cele 100 bile din urnă. Sînt posibile deci 100 de cazuri. La extracția a doua, bila extrasă anterior revenind în urnă, fiecare din cele 100 de cazuri din prima extracție se pot asocia cu alte 100 de cazuri. Prin urmare, după două extracții sînt egal posibile 100^2 cazuri. În mod analog, după trei extracții sînt posibile 100^3 cazuri, iar după patru exerciții, 100^4 cazuri, egal posibile.

Numărul cazurilor favorabile îl putem determina, folosind notația: cazul de apariție a unei bile albe cu A și cel de apariție a unei bile negre cu N.

Succesiunile favorabile apariției a două bile albe în cele patru extracții sînt următoarele șase:

AANN; ANAN; ANNA; NAAN; NANA; NNAA.

Dacă luăm, de pildă, prima din aceste succesiuni, putem observa că ea se poate obține în $80^2 \cdot 20^2$ moduri diferite.

Este evident că dacă în prima extracție s-a realizat evenimentul elementar de apariție a unei bile albe, el are 80 de variante diferite întrucît în urnă sînt 80 de bile albe; în extracția a doua fiecare din cele 80 de variante ale bilei albe obținute în prima extracție se poate asocia cu 80 de variante ale bilei albe obținute în extracția a doua. Așadar evenimentul ca în prima și a doua extracție să apară cîte o bilă albă se poate realiza în $80 \cdot 80 = 80^2 = 6\,400$ moduri diferite. Fiecăruia din aceste moduri i se poate asocia un număr de $20 \cdot 20 = 20^2 = 400$ cazuri de apariție a bilei negre în extracțiile a treia și a patra. În total, evenimentul com-

plex al succesiunii AANN se poate realiza în $6\,400 \cdot 400 = 80^2 \cdot 20^2 = 2\,560\,000$ moduri.

Același raționament este valabil și pentru celelalte cinci succesiuni care sînt de același tip.

Vom avea deci un număr de $6 \cdot 80^2 \cdot 20^2$ cazuri favorabile pentru ca în patru extracții succesive să obținem două bile albe.

Avînd calculat numărul cazurilor egal posibile și numărul cazurilor favorabile, probabilitatea căutată P va fi egală cu

$$P = \frac{6 \cdot 80 \cdot 80 \cdot 20 \cdot 20}{100\,000\,000} = \frac{96}{625}.$$

Vom vedea că această problemă se poate rezolva mai simplu cunoscînd anumite reguli de calcul al probabilităților.

*

Pentru calculul practic al probabilităților se folosesc două reguli importante care se desprind cu ușurință din schema axiomelor lui Kolmogorov.

A. *Regula adunării probabilităților.* În cazul unei reuniuni de evenimente incompatibile, probabilitatea evenimentului reunit este egală cu suma probabilităților de a se realiza fiecare eveniment în parte (axioma IV).

Pentru ilustrarea concretă a acestei reguli ne poate servi exemplul 5 de mai sus. Referitor la acest exemplu ne putem întreba care este probabilitatea ca să obținem una din sumele 2,3 sau 4 ale punctelor? Este evident că evenimentul de a obține una din aceste sume exclude posibilitatea de a se obține celelalte două. De exemplu, dacă obținem suma 4, se exclude posibilitatea de a fi obținute, simultan, sumele 2 sau 3; evenimentele sînt deci, *incompatibile*. Potrivit regulii enunțate, probabilitatea evenimentului reunit, adică probabilitatea realizării fie a sumei 2, fie a sumei 3, fie a sumei 4, este egală cu suma probabilităților de a se realiza fiecare eveniment în parte.

Ținînd seama de notațiile folosite în capitolul I, vom scrie

$$P[(2) \cup (3) \cup (4)] = P(2, 3, 4) = P(2) + P(3) + P(4) =$$

$$= \frac{1}{36} + \frac{2}{36} + \frac{3}{36} = \frac{6}{36}.$$

Referitor la același exemplu ne putem pune întrebarea care este probabilitatea de a obține o sumă oarecare?

Această probabilitate o putem calcula în două moduri:

a) Vom calcula probabilitatea reuniunii de evenimente elementare (în cazul nostru evenimentele elementare sînt cele 11 sume posibile,

întrucît sînt incompatibile între ele). Deci, vom avea

$$P\,[(2) \cup (3) \cup \ldots \cup (11) \cup (12)] = P(2, 3, \ldots, 10, 11, 12) =$$
$$= P(2) + P(3) + \ldots + P(11) + P(12) =$$
$$= \frac{1}{36} + \frac{2}{36} + \frac{3}{36} + \frac{4}{36} + \frac{5}{36} + \frac{6}{36} +$$
$$+ \frac{5}{36} + \frac{4}{36} + \frac{3}{36} + \frac{2}{36} + \frac{1}{36} = \frac{36}{36} = 1.$$

b) Vom calcula probabilitatea unei reuniuni de evenimente incompatibile complementare. De exemplu, probabilitatea reuniunii evenimentelor (2, 3, 4, 5, 6, 7, 8) și (9, 10, 11, 12). Ne putem convinge ușor că această probabilitate este egală cu 1.

Prin urmare, procedînd într-un fel sau altul am obținut un *eveniment sigur* din același cîmp de evenimente al experienței noastre. Într-adevăr, aruncînd de două ori un zar este sigur că vom obține una din cele 11 sume posibile ale punctelor ce se obțin din prima și a doua aruncare.

Subliniem că regula adunării probabilităților poate fi aplicată numai în cazul evenimentelor incompatibile. Nerespectarea acestei condiții poate duce la rezultate absurde.

Să presupunem astfel că în exemplul 6 de mai sus, am numerotat bilele albe cu 1—80 inclusiv, iar pe cele negre am trecut numerele de la 81 la 100. În aceste condiții să luăm în considerare evenimentul E de a scoate din urnă fie o bilă albă, fie o bilă pe care se găsește scris un număr cu soț. Evenimentul E reprezintă o reuniune a evenimentelor E_1 (evenimentul de a fi scoasă o bilă albă) și E_2 (evenimentul de a fi scoasă o bilă cu soț).

Probabilitatea ca să se verifice E_1 este

$$P_1 = \frac{80}{100},$$

iar probabilitatea ca să se verifice evenimentul E_2 este

$$P_2 = \frac{50}{100}$$

întrucît numărul cazurilor favorabile este egal cu al numărului de bile care au trecute pe ele un număr cu soț, adică 50, pe cînd numărul cazurilor posibile, reprezentînd totalitatea bilelor din urnă, este 100.

Dacă vom aplica regula adunării probabilităților fără ca să examinăm dacă evenimentele E_1 și E_2 sînt incompatibile, vom obține pentru probabilitatea p de a se produce evenimentul E valoarea :

$$p = p_1 + p_2 = \frac{80}{100} + \frac{50}{100} = \frac{130}{100},$$

adică un număr mai mare decît 1 ceea ce este imposibil, dacă ținem seama de dubla inegalitate din (I, 2, 2).

Eroarea provine din faptul că evenimentele E_1 și E_2 nu se exclud unul pe altul, ele sînt compatibile și de aceea regula adunării probabilităților nu este aplicabilă.

Raționamentul exact pentru aflarea probabilității evenimentului E este următorul:

Evenimentul E este o reuniune a următoarelor evenimente *incompatibile*:

a) scoaterea unei bile albe purtînd pe ea un număr fără soț;
b) scoaterea unei bile albe purtînd pe ea un număr cu soț;
c) scoaterea unei bile negre purtînd pe ea un număr cu soț.

Dacă însemnăm prin E_1, E_2, E_3 aceste evenimente din cîmpul de evenimente al experienței noastre și ne propunem să determinăm probabilitățile lor de a se produce, pe care le însemnăm respectiv prin p_1, p_2, p_3, vom avea

$$p_1 = \frac{40}{100} \text{ și } p_2 = \frac{40}{100}$$

întrucît din 80 de bile albe din urna cu 100 de bile, jumătate poartă numerele cu soț și jumătate fără soț iar

$$p_3 = \frac{10}{100}$$

întrucît din cele 20 de bile negre avem 10 care poartă pe ele un număr cu soț.

Prin urmare, aplicînd regula adunării probabilităților, probabilitatea P a evenimentului E este

$$P = p_1 + p_2 + p_3 = \frac{40}{100} + \frac{40}{100} + \frac{10}{100} = \frac{90}{100} = \frac{9}{10},$$

ceea ce înseamnă că în nouă din zece extracții, neschimbînd structura urnei, avem probabilitatea teoretică de a obține sau o bilă albă sau o bilă care poartă pe ea un număr cu soț, indiferent de culoare.

Pentru aplicarea regulei adunării probabilităților într-un caz mai general putem lua următorul exemplu:

EXEMPLUL 7. Probabilitatea ca să se verifice evenimentul E_1 este p_1; probabilitatea ca să se verifice evenimentul E_2 este p_2; iar probabilitatea ca să se verifice amîndouă evenimentele — evenimentul E_3 — este p_3. Care este probabilitatea ca să se verifice cel puțin unul din ele?

Vom observa că evenimentul reunit E a cărui probabilitate o căutăm, se realizează printr-unul din următoarele trei evenimente:

— evenimentul E_1 *se întîmplă singur*;
— evenimentul E_2 *se întîmplă singur*;
— evenimentele E_1 și E_2 se întîmplă împreună, E_3.

Să însemnăm prin P_1 probabilitatea ca să se verifice singur primul din evenimente, iar prin P_2 probabilitatea ca să se verifice singur al doilea.

În baza regulei de adunare a probabilităților — evenimentele fiind incompatibile — probabilitatea căutată, pe care o însemnăm cu P, este

$$P = P_1 + P_2 + p_3, \qquad \text{(II,2.1)}$$

deoarece evenimentul E_3 se realizează cu probabilitatea p_3.

Mai observăm însă că evenimentul E_1 se realizează la rîndul lui prin două evenimente deosebite:

— E_1 se produce singur;

— E_1 se produce împreună cu E_2 — realizîndu-se E_3.

Probabilitatea corespunzătoare primului eveniment component al lui E_1 este P_1, iar celui de-al doilea, p_3. Cum probabilitatea de realizare a evenimentului E_1 este p_1, rezultă în baza regulei de adunare a probabilităților că:

$$p_1 = P_1 + p_3$$

sau

$$P_1 = p_1 - p_3, \qquad \text{(II,2.2)}$$

ceea ce înseamnă că probabilitatea P_1 a unui eveniment E_1 de a se realiza *singur* este egală cu probabilitatea p_1 a acelui eveniment de a se realiza *fie singur, fie împreună* cu un alt eveniment E_2 compatibil, mai puțin probabilitatea p_3 de a se realiza împreună cu evenimentul E_2.

În mod analog se demonstrează că

$$P_2 = p_2 - p_3. \qquad \text{(II,2.3)}$$

Înlocuind în egalitatea (II, 2.1) valorile lui P_1 și P_2 prin valorile găsite în (II, 2.2) și (II, 2.3), rezultă

$$P = (p_1 - p_3) + (p_2 - p_3) + p_3 = p_1 + p_2 - p_3. \qquad \text{(II,2.4)}$$

Este de observat că formula găsită generalizează regula adunării probabilităților pentru două evenimente oarecare. În cazul particular cînd evenimentele se exclud între ele avem $p_3 = 0$ și deci se regăsește regula adunării probabilităților.

B. *Regula probabilităților compuse* (*înmulțirii probabilităților*). Pentru lămurirea acestei reguli trebuie să precizăm în prealabil sensul a două categorii de evenimente și anume:

— *evenimente independente*;

— *evenimente dependente.*

Evenimentele E_1, E_2, ..., E_m sînt independente dacă probabilitatea fiecărui eveniment de acest fel nu este prin nimic înrîurită de evenimentele petrecute anterior.

Folosim această definiție, utilă pentru aplicațiile practice ale calculului probabilităților. Ea este suficientă în cazul cînd se poate folosi definiția clasică a probabilității, adică în cazul cînd evenimentele sînt în număr finit.

De exemplu, faptul că în exemplul 6 considerat mai sus s-a pus condiția ca după fiecare extracție să se pună bila înapoi în urnă a făcut ca

probabilitatea de a extrage o bilă albă (la fel ca și probabilitatea de a extrage o bilă neagră) să rămînă constantă la fiecare extracție; evenimentele de a extrage o bilă albă la prima extracție, la a doua sau la oricare alta, sînt *evenimente independente.*

Dacă însă, în același exemplu, s-ar face extracții succesive fără a se pune bila înapoi în urnă, probabilitatea de a se realiza un eveniment oarecare — de exemplu, extragerea unei bile albe la a treia extracție — ar fi influențată de rezultatele obținute în primele două extracții.

Astfel, dacă în primele două extracții am obținut bile albe, în urnă găsindu-se 80 de bile albe și 20 de bile negre, probabilitatea de a extrage o bilă albă și în extracția a treia ar fi

$$p = \frac{78}{98}$$

iar probabilitatea complementară de a se extrage o bilă neagră la extracția a treia ar fi

$$q = \frac{20}{98}$$

întrucît numărul bilelor negre rămase în urnă după a doua extracție este egal cu numărul lor inițial, iar numărul total al cazurilor egal posibile s-a micșorat cu doi.

Verificarea faptului că realizarea unor evenimente anterioare nu influențează asupra probabilității ca un nou eveniment să se producă nu este întotdeauna așa de simplă ca în exemplul precedent. În statistică stabilirea independenței sau a gradului de dependență a două sau mai multe fenomene a dat naștere unei teorii, care face apel la cunoștințe statistice avansate, numită teoria corelației. După cum vom vedea la capitolul XI stabilirea independenței a două sau mai multe evenimente conduce la calcule complicate și nu este întotdeauna posibilă. Se impune de aceea o definiție simplă, precisă, matematică, a independenței a două evenimente.

Să considerăm două evenimente A și B. Prin $P(A \cap B)$ notăm probabilitatea ca să se întîmple amîndouă evenimentele împreună, prin $P(A)$ probabilitatea ca să se întîmple A, prin $P(B)$ probabilitatea ca să se întîmple B. Notăm

$$P_A(B) = \frac{P(A \cap B)}{P(A)}, \qquad P_B(A) = \frac{P(A \cap B)}{P(B)}.$$

Se poate arăta ușor că numerele $P_A(B)$ și $P_B(A)$, definite cu ajutorul evenimentelor A și B, îndeplinesc toate axiomele unei probabilități. Ele sînt deci niște probabilități. Prin definiție, vom spune că $P_A(B)$ reprezintă probabilitatea evenimentului B, condiționată de A, iar $P_B(A)$ reprezintă probabilitatea evenimentului A, condiționată de B.

Dacă avem

$$P(A \cap B) = P(A) \cdot P(B)$$

vom spune, prin definiție, că evenimentele A și B sînt independente între ele.

În acest caz avem

$$P_A(B) = \frac{P(A \cap B)}{P(A)} = \frac{P(A) \cdot P(B)}{P(A)} = P(B),$$

$$P_B(A) = \frac{P(A \cap B)}{P(B)} = \frac{P(A) \cdot P(B)}{P(B)} = P(A).$$

Scanned with OKEN Scanner

Egalitățile:

$$P_A(B) = P(B) \text{ și } P_B(A) = P(A)$$

arată că condiționarea lui B de A și a lui A de B, nu are nici o influență asupra probabilităților $P(A)$ și $P(B)$. Evenimentele A și B sînt independente.

Dacă avem n evenimente $A_1, A_2, \ldots, A_n$, ele sînt independente dacă pentru oricare subșir $1 \leqslant i_1 < i_2 < \ldots < i_\gamma \leqslant n$, avem

$$P(A_{i_1} \cap A_{i_2} \cap \ldots \cap A_{i_\gamma}) = P(A_{i_1})\ P(A_{i_2}) \ldots P(A_{i_\gamma}).$$

a) *Regula probabilităților compuse în cazul evenimentelor independente.* În cazul cînd numărul evenimentelor este finit, aplicîndu-se definiția clasică a probabilității, se verifică ușor că *probabilitatea ca un șir de evenimente independente să se întîmple împreună este egală cu produsul probabilităților ca fiecare din ele să se întîmple separat.*

Să considerăm șirul de evenimente independente $E_1, E_2, \ldots, E_m$ și să notăm cu $N_1, N_2, \ldots, N_m$ numărul cazurilor posibile corespunzătoare fiecăruia dintre ele. De asemenea, fie $n_1, n_2, \ldots, n_m$ numărul respectiv al cazurilor favorabile.

Probabilitățile simple corespunzătoare vor fi

$$p_1 = \frac{n_1}{N_1}, \qquad p_2 = \frac{n_2}{N_2}, \ldots, p_m = \frac{n_m}{N_m}.$$

Cu ajutorul acestor probabilități simple ne propunem să determinăm *probabilitatea compusă de a realiza toate evenimentele* $E_1, E_2, \ldots, E_m$ *împreună.*

Să considerăm, în primul rînd, două evenimente E_1 și E_2. Pentru E_1 avem n_1 cazuri favorabile, iar pentru E_2 avem n_2 asemenea cazuri. Numărul cazurilor favorabile pentru ca evenimentele E_1 și E_2 să se verifice împreună este, evident, egal cu $n_1 \cdot n_2$; într-adevăr, la fiecare caz favorabil prin care se realizează E_1 se poate asocia cîte unul din cele n_2 cazuri favorabile pentru E_2. Vom avea în total $n_1 \cdot n_2$ cazuri favorabile pentru realizarea împreună a evenimentelor E_1 și E_2.

Dacă considerăm trei evenimente E_1, E_2, E_3 — pe baza aceluiași raționament — vom avea $n_1 \cdot n_2 \cdot n_3$ cazuri favorabile pentru realizarea împreună a evenimentelor E_1, E_2 și E_3 (fiecare caz favorabil din numărul de $n_2 \cdot n_3$ evenimente favorabile realizării simultane a evenimentelor E_2 și E_3, se asociază cu fiecare caz favorabil din numărul de n_1 de cazuri favorabile realizării evenimentului E_1).

Raționamentul făcut este general.

Putem deci afirma că numărul cazurilor favorabile pentru ca evenimentele independente $E_1, E_2, \ldots, E_m$ să se verifice împreună este egal cu produsul

$$n_1 \cdot n_2 \ldots n_m.$$

În mod analog se arată că numărul cazurilor posibile este egal cu $N_1 \cdot N_2 \ldots N_m$.

Scanned with OKEN Scanner

Drept consecință, probabilitatea P ca aceste evenimente să se realizeze împreună este

$$P = \frac{n_1 \cdot n_2 \dots n_m}{N_1 \cdot N_2 \dots N_m} = \frac{n_1}{N_1} \cdot \frac{n_2}{N_2} \dots \frac{n_m}{N_m} = p_1 \cdot p_2 \dots p_m,$$

relație care demonstrează regula probabilităților compuse enunțată mai sus.

Să considerăm următorul exemplu:

EXEMPLUL 8. Se aruncă patru zaruri perfect construite. Care este probabilitatea ca pe fiecare zar să iasă fața cu șase puncte?

Observăm că probabilitatea ca să iasă fața cu șase puncte pe oricare din zaruri este $\frac{1}{6}$, deoarece din cele șase fețe ale fiecărui zar, avem una care este favorabilă.

Mai observăm că evenimentele elementare ca să nu apară pe fiecare zar fața cu șase puncte sînt independente și că deci, aplicînd teorema probabilităților compuse, probabilitatea ca aceste evenimente să se producă împreună va fi:

$$\frac{1}{6} \cdot \frac{1}{6} \cdot \frac{1}{6} \cdot \frac{1}{6} = \frac{1}{1296};$$

un asemenea eveniment este, evident, rar, deoarece teoretic el poate apare odată la 1296 aruncări de patru zaruri.

Dacă revenim acum din nou la exemplul 6 considerat mai sus, vom putea calcula, mult mai simplu, probabilitatea ca din cele patru bile extrase două să fie albe, prin folosirea regulilor de calcul examinate.

Observăm că probabilitatea de a extrage o bilă albă din urna dată (cu 80 de bile albe și 20 de bile negre) este de $\frac{80}{100} = \frac{4}{5}$, iar probabilitatea de a extrage o bilă neagră este de $\frac{20}{100} = \frac{1}{5}$; aceste probabilități rămîn constante pentru fiecare extracție, deoarece după fiecare extracție bila se pune înapoi în urnă (vom vedea mai departe că sîntem în cazul urnei Bernoulli).

Mai observăm că cele șase succesiuni favorabile (AANN; ANAN;...) sînt evenimente incompatibile, neputînd avea loc simultan. Or, pe baza axiomei a patra a lui Kolmogorov (regula adunării probabilităților), probabilitatea ca să se verifice una din ele, oricare, este egală cu suma probabilităților fiecăreia în parte.

Să determinăm, deci, probabilitatea uneia din succesiuni (care este aceeași pentru fiecare în parte, succesiunile fiind de același tip) de pildă, probabilitatea succesiunii ANAN.

Observăm că apariția fiecărui eveniment elementar din această succesiune constituie un eveniment independent și în consecință, proba-

bilitatea ca cele patru evenimente să aibă loc succesiv împreună este egală cu produsul probabilităților fiecărui eveniment elementar, adică:

$$\frac{4}{5}\cdot\frac{1}{5}\cdot\frac{4}{5}\cdot\frac{1}{5}=\frac{16}{625}.$$

Însumînd de șase ori probabilitatea fiecărei succesiuni, obținem $\frac{96}{625}$, probabilitatea căutată.

b) *Regula probabilităților compuse în cazul evenimentelor dependente.* Această regulă, mult mai generală decît regula probabilităților compuse în cazul evenimentelor independente, ne ajută să determinăm probabilitatea ca m evenimente *dependente* $E_1, E_2, \ldots, E_m$ să se producă împreună.

Ea poate fi formulată în felul următor:

Fie $E_1, E_2, \ldots, E_m$ un șir de evenimente dependente. Probabilitatea ca să se realizeze succesiv toate aceste evenimente împreună este egală cu produsul probabilităților de a se realiza fiecare eveniment după ce evenimentele care-l preced s-au realizat.

Pentru înțelegerea intuitivă a acestei reguli se folosește, în general, același raționament ca la demonstrarea regulii în cazul evenimentelor independente, ținîndu-se seama însă că realizarea evenimentului E_1 schimbă condițiile în care se realizează evenimentul E_2, iar realizarea evenimentelor E_1 și E_2 schimbă condițiile în care se realizează evenimentul E_3 și așa mai departe.

Schematic acest raționament se poate prezenta astfel:

Dacă însemnăm prin:

n_1, numărul cazurilor favorabile pentru E_1;
n_2, numărul cazurilor favorabile pentru E_2 după realizarea lui E_1;
n_3, numărul cazurilor favorabile pentru E_3 după realizarea lui E_1, E_2;
. . . . . .
n_m, numărul cazurilor favorabile pentru E_m după realizarea lui E_1, E_2,..,E_{m-1},

atunci numărul cazurilor favorabile pentru realizarea împreună a evenimentelor E_1, E_2,..,E_m este dat de produsul

$$n_1 \cdot n_2 \ldots n_m .$$

În același mod se arată că dacă însemnăm prin:

N_1, numărul cazurilor posibile pentru E_1;
N_2, numărul cazurilor posibile pentru E_2 după realizarea lui E_1;
N_3, numărul cazurilor posibile pentru E_3 după realizarea lui E_1, E_2;
. . . . . .
N_m, numărul cazurilor posibile pentru E_m după realizarea lui E_1, E_2,...,E_{m-1},

numărul cazurilor posibile pentru realizarea împreună a evenimentelor E_1, E_2,...,E_m este dat de produsul

$$N_1 \cdot N_2 \cdots N_m .$$

Așadar, probabilitatea ca să se realizeze toate evenimentele E_1, E_2,..., E_m împreună este egală cu

$$P=\frac{n_1\cdot n_2\cdots n_m}{N_1\cdot N_2\cdots N_m}=\frac{n_1}{N_1}\cdot\frac{n_2}{N_2}\cdots\frac{n_m}{N_m}=p_1\cdot p_2\cdots p_m$$

în care valorile $p_1, p_2, \ldots, p_m$ au o semnificație precisă și anume:

p_1 reprezintă probabilitatea de a se realiza E_1;
p_2 reprezintă probabilitatea de a se realiza E_2 după ce s-a realizat E_1;
p_3 reprezintă probabilitatea de a se realiza E_3 după ce s-a realizat E_1, E_2;
. . . . . .
p_m reprezintă probabilitatea de a se realiza E_m după ce s-a realizat $E_1, E_2, \ldots, E_{m-1}$

Să considerăm următorul exemplu:

EXEMPLUL 9. Într-o urnă sînt 52 de bile din care 48 sînt bile albe iar 4 bile roșii. Se fac succesiv două extracții din urnă, fără ca să se pună bila extrasă înapoi în urnă. Să se calculeze probabilitatea ca ambele bile extrase să fie roșii.

Vom avea două evenimente și anume E_1 care reprezintă scoaterea unei bile roșii în prima extracție și E_2, de a scoate o bilă roșie și în a doua extracție. Aceste evenimente, evident, nu mai sînt independente, deoarece dacă în prima extracție am obținut o bilă roșie mai rămîn în urnă numai 51 de bile din care numai 3 bile roșii. Probabilitatea de a se verifică evenimentul E_2, adică de a fi scoasă o bilă roșie în a doua extracție, este deci $\frac{3}{51}$.

Probabilitatea de a se realiza evenimentul E_2 se schimbă, prin urmare, după modul cum s-a realizat E_1; sîntem în prezența a două evenimente *dependente*.

Pentru determinarea probabilității de a se realiza ambele evenimente putem aplica regula probabilităților compuse în cazul evenimentelor dependente.

Vom avea probabilitatea evenimentului E_1 egală cu $\frac{4}{52}$, iar a evenimentului E_2, după ce s-a realizat E_1, egală, după cum am văzut, cu $\frac{3}{51}$. Probabilitatea căutată este egală, prin urmare, cu produsul:

$$\frac{4}{52} \cdot \frac{3}{51} = \frac{1}{13} \cdot \frac{1}{17} = \frac{1}{221}.$$

Să mai considerăm un exemplu în care putem aplica regulile de calcul al probabilităților:

EXEMPLUL 10. Dintr-o urnă cu 75 de bile roșii și 25 de bile albe se fac succesiv trei extracții, fără ca să se pună bila înapoi în urnă. Să se calculeze probabilitatea ca două din cele trei bile extrase să fie albe.

Problema este de același tip ca cea din exemplul precedent.

Însemnăm prin A extragerea unei bile albe și prin R extragerea unei bile roșii.

Vom avea de considerat succesiunile de bile:

RAA, ARA, AAR,

unde sînt arătate culorile bilelor în ordinea extragerii lor.

Aceste succesiuni sînt trei evenimente ce se exclud reciproc — evenimente incompatibile — iar reuniunea lor constituie un alt eveniment a cărui probabilitate o căutăm.

Prin urmare, va fi suficient să calculăm pentru fiecare din ele probabilitatea respectivă de a se realiza și apoi, aplicînd regula adunării probabilităților, să le însumăm.

Să considerăm mai întîi succesiunea RAA care înseamnă că în prima extracție a ieșit bila roșie, iar în celelalte două cîte una albă. Este evident că extracțiile de bile nu mai formează evenimente independente între ele, deoarece culoarea bilelor obținută în extracțiile precedente influențează asupra probabilității de a scoate o bilă de culoare dată într-o extracție de rang anumit.

Într-adevăr, dacă am scos o bilă roșie din prima extracție, în urnă rămîn în total 99 bile din care 74 bile roșii; probabilitatea de a fi scoasă o bilă roșie în extracția a doua va fi, deci, $\frac{74}{99}$. Dacă, însă, din prima extracție a ieșit o bilă albă, numărul total al bilelor din urnă va fi tot de 99 însă numărul bilelor roșii este 75; probabilitatea ca să scoatem o bilă roșie din extracția a doua este de data aceasta $\frac{75}{99}$. În extracția a doua, deci, probabilitatea de a scoate o bilă roșie este $\frac{74}{99}$ sau $\frac{75}{99}$, după cum la prima extracție a ieșit o bilă roșie sau o bilă albă.

Pentru calculul probabilității succesiunii de bile RAA, vom aplica regula înmulțirii probabilităților în cazul evenimentelor dependente.

Probabilitatea ca să scoatem din prima extracție o bilă roșie este de $\frac{75}{100}$.

Probabilitatea ca să scoatem în a doua extracție o bilă albă după ce în prima extracție am obținut o bilă roșie, este $\frac{25}{99}$.

În sfîrșit, probabilitatea ca în a treia extracție să mai scoatem o bilă albă, după ce în prima extracție s-a scos o bilă roșie, iar în a doua o bilă albă, este $\frac{24}{98}$, deoarece numărul bilelor rămase în urnă este 98 din care numai 24 albe.

Probabilitatea succesiunii de bile RAA este prin urmare produsul

$$\frac{75}{100} \cdot \frac{25}{99} \cdot \frac{24}{98}.$$

În mod analog se arată că probabilitatea succesiunii ARA este

$$\frac{25}{100} \cdot \frac{75}{99} \cdot \frac{24}{98}$$

iar a succesiunii AAR este

$$\frac{25}{100} \cdot \frac{24}{99} \cdot \frac{75}{98}.$$

Probabilitatea căutată se obține — după cum am mai arătat — prin adunarea probabilităților celor trei succesiuni.

Avem deci

$$\frac{75}{100} \cdot \frac{25}{99} \cdot \frac{24}{98} + \frac{25}{100} \cdot \frac{75}{90} \cdot \frac{24}{98} + \frac{25}{100} \cdot \frac{24}{99} \cdot \frac{75}{98} = 3 \cdot \frac{75 \cdot 25 \cdot 24}{100 \cdot 99 \cdot 98} = \frac{75}{539}.$$

Acest raport ne dă valoarea probabilității căutate.

Capitolul III

Scheme clasice din teoria probabilităţilor

1. Schema urnei Bernoulli

Observînd în natură anumite fenomene care se repetă de nenumărate ori în condiţii identice, J. Bernoulli şi-a propus să construiască o schemă probabilistică cu ajutorul căreia să poată studia, din punct de vedere probabilistic, asemenea fenomene.

El a considerat următoarea schemă generală:

O urnă conţine bile de diferite culori, în proporţii date. Se cere probabilitatea ca din n bile extrase, punînd de fiecare dată bila extrasă înapoi în urnă, să obţinem n_1 bile de o culoare, n_2 bile de altă culoare etc.

Urna Bernoulli cu două stări. Exemplul 6 considerat în capitolul II este un caz tipic de schemă probabilistică a urnei Bernoulli, cu *două stări* (urna conţine bile de două culori), avînd o largă aplicaţie în domeniul fenomenelor care se repetă de numeroase ori în condiţii identice.

Este vorba prin urmare de o urnă în care se află A bile albe şi B bile negre. Din această urnă se efectuează un şir de extracţii succesive, punînd după fiecare extracţie bila înapoi în urnă. În modul acesta compoziţia urnei rămîne aceeaşi şi probabilitatea de a scoate o bilă albă rămîne constantă în toate extracţiile şi anume

$$p = \frac{A}{A + B}$$

iar probabilitatea complementară (contrarie) de a fi scoasă o bilă neagră este

$$q = \frac{B}{A + B}.$$

Fie E_1, E_2, E_3, ... evenimentele de a scoate o bilă albă respectiv în extracţia întîi, a doua, ... Este evident că aceste evenimente sînt inde-

pendente între ele și prin urmare, oricare ar fi rezultatul extracției precedente, probabilitatea de a obține o bilă albă în extracția următoare este întotdeauna aceeași; probabilitățile $p_1, p_2, p_3, \ldots$ rămîn mereu aceleași.

Dacă facem, în general, o serie de n probe independente care dau naștere la două evenimente cu probabilitățile respective p și q în fiecare probă, în baza axiomei a patra a lui Kolmogorov, avem, evident,

$$p + q = \frac{A}{A+B} + \frac{B}{A+B} = 1,$$

deoarece evenimentul reunit de a se realiza unul din evenimentele elementare, sau A sau B, este un eveniment sigur.

Să considerăm un exemplu:

EXEMPLUL 11. Se aruncă de opt ori o monedă. Se cere probabilitatea de a obține stema de patru ori.

Moneda poate fi asemănată cu o urnă în care există două bile de culori diferite — sau în general — bile de două culori, în proporție egală. Probabilitatea de a extrage o bilă de culoare dată egală cu $\frac{1}{2}$.

În speță, probabilitatea de a obține fața cu stemă este egală cu $\frac{1}{2} = p$; tot cu $\frac{1}{2}$ este egală și probabilitatea complementară de a se obține cealaltă față (marca).

Este clar că succesiunile favorabile evenimentului de a se obține de patru ori fața cu stemă în opt aruncări sînt de forma: MMBBMMBB; MBBBMBMM ... și că ele sînt în număr egal cu numărul combinărilor ce se poate face cu opt obiecte luate cîte patru, întrucît interesează posibilitatea teoretică de apariție distinctă a unei fețe și nu ordinea în care va apare.

Avem deci $C_8^4 = 70$ asemenea succesiuni. Fiecare din aceste succesiuni are aceeași probabilitate de a se verifica, a cărui valoare este dată de produsele

$$\left(\frac{1}{2}\right)^4 \cdot \left(\frac{1}{2}\right)^4$$

întrucît sîntem în cazul evenimentelor independente iar probabilitatea de a se verifica succesiv împreună cele opt evenimente elementare, este egală cu produsul probabilităților ca fiecare dintre ele să se verifice separat.

Însumînd probabilitățile tuturor celor 70 de succesiuni, vom determina probabilitatea căutată de a fi verificat evenimentul apariției de patru ori a stemei. Dacă notăm această probabilitate cu $P_{(4,4)}$, avem

$$P_{(4,4)} = C_8^4 \left(\frac{1}{2}\right)^4 \cdot \left(\frac{1}{2}\right)^4 = \frac{8!}{4!\,4!} \cdot \frac{1}{16} \cdot \frac{1}{16} = 70 \cdot \frac{1}{256} = \frac{35}{128},$$

Acest exemplu ne sugerează — pentru cazul urnei lui Bernoulli cu două stări — următoarea formulă a probabilității de a se verifica

evenimentul considerat de un anumit număr k de ori, în n probe succesive :

$$P_{(k,n-k)} = \frac{n!}{k!\,(n-k)!} \cdot p^k \cdot q^{n-k}, \qquad \text{(III,1.1)}$$

In care :

$P_{(k,n-k)}$ reprezintă probabilitatea ca dintr-o urnă Bernoulli cu două stări (cu bile de două culori, de pildă, albe și negre) în n extracții succesive, o bilă de anumită culoare (evenimentul E) să apară de k ori, iar bila de cealaltă culoare (evenimentul complementar E) să apară, respectiv, de $n-k$ ori ;

n, numărul extracțiilor independente ce se efectuează ;

k, numărul cazurilor în care apare bila de o anumită culoare ($k = 0, 1, 2, .., n$) ;

p, probabilitatea evenimentului elementar ca într-o extracție să se verifice apariția bilei de culoare dată ;

$q = 1-p$, idem, — ca într-o extracție să se verifice apariția bilei de cealaltă culoare.

În continuare, ori de cîte ori vom considera o urnă Bernoulli cu bile albe și negre, vom folosi formula (III,1.1) într-o formă mai prescurtată, în care vom nota cu $P_{n(\alpha)}$ probabilitatea ca în n extracții succesive bila albă să apară de α ori ($\alpha = 0, 1, 2, \ldots, n$) :

$$P_{n(\alpha)} = C_n^\alpha p^\alpha q^{n-\alpha}, \qquad \text{(III,1.2)}$$

Se vede acum că această formulă se poate aplica și la rezolvarea problemei pusă în exemplul 6. Vom avea $\alpha = 2$; $n = 4$; $p = \frac{4}{5}$; $q = \frac{1}{5}$; probabilitatea căutată este

$$P_{4(2)} = C_4^2 \left(\frac{4}{5}\right)^2 \cdot \left(\frac{1}{5}\right)^2 = 6 \cdot \frac{16}{25} \cdot \frac{1}{25} = \frac{96}{625}.$$

2. Schema multinomială

Dacă în cazul urnei Bernoulli, pe care l-am tratat, în loc de bile cu două culori avem bile de m culori și extracțiile succesive din urnă le facem în același mod (punînd de fiecare dată bila înapoi în urnă), atunci, într-o extracție din această urnă putem obține unul din evenimentele :

$$E_1, E_2, \ldots, E_m$$

incompatibile între ele, constînd în scoaterea unei bile de o anumită culoare. Aceasta înseamnă că dacă vom nota cu p_i probabilitatea evenimentului elementar E_i (în care $i = 1, 2, \ldots, n$), adică p_1, probabilitatea ca să obținem bila de culoarea întîi, p_2, probabilitatea să apară bila de culoare a doua etc., avem :

$$p_1 + p_2 + \ldots + p_m = 1.$$

Ne propunem să calculăm probabilitatea $P_n(\lambda_1, \lambda_2, \ldots, \lambda_m)$ ca în n extracții evenimentul E_1 (de a fi extrasă bila de culoare întîi) să se verifice de λ_1 ori, evenimentul E_2 (de a fi extrasă bila de culoarea a doua)

să se verifice de λ_2 ori, ..., evenimentul E_m (de a fi extrasă bila de culoarea m) să se verifice de λ_m ori, unde

$$\lambda_1 + \lambda_2 + \ldots + \lambda_m = n.$$

Există pentru probabilitatea căutată următoarea formulă:

$$P_{n(\lambda_1,\lambda_2,\ldots,\lambda_m)} = \frac{n!}{\lambda_1!\,\lambda_2!\ldots\lambda_m!} \cdot p_1^{\lambda_1} p_2^{\lambda_2} \ldots p_m^{\lambda_m} \qquad \text{(III,2.1)}$$

în care expresia $\frac{n!}{\lambda_1!\,\lambda_2!\ldots\lambda_m!}$ este cunoscuta formulă a permutărilor repetate de n elemente din care λ_1, respectiv λ_2, ..., λ_m sînt egale între ele.

EXEMPLUL 12. Un magazin vinde în cursul unei săptămîni 2 600 bucăți dintr-o anumită marfă, primită de la fabricile A, B, C în următoarele cantități:

De la fabrica A	3 000	bucăți
De la fabrica B	2 600	„
De la fabrica C	4 200	„
Total de la fabricile A, B, C. . .	9 800	bucăți.

Marfa fiind de aceeași calitate, să se determine probabilitatea ca din cantitatea vîndută săptămînal, 820 bucăți să fie de la fabrica A, 500 bucăți de la fabrica B și restul de 1 280 bucăți de la fabrica C.

În această problemă există evenimentele:

E_1 de a vinde o unitate din marfa procurată de la fabrica A;
E_2 de a vinde o unitate din marfa procurată de la fabrica B;
E_3 de a vinde o unitate din marfa procurată de la fabrica C.

Aceste evenimente sînt incompatibile între ele. Admițînd că cele 9 800 de bucăți sînt amestecate și clienții nu manifestă nici o preferință, dată fiind aceeași calitate a mărfurilor, avem probabilitățile:

$$P_1 = \frac{3\,000}{9\,800} = \frac{15}{49},$$

$$P_2 = \frac{2\,600}{9\,800} = \frac{13}{49},$$

$$P_3 = \frac{4\,200}{9\,800} = \frac{21}{49},$$

ale evenimentelor respective E_1, E_2 și E_3.

Observăm că sîntem în cazul unei urne Bernoulli cu trei stări, în care: numărul stărilor $m = 3$; numărul probelor (vînzărilor) într-o săptămînă $n = 2\,600$, iar numărul cazurilor de vînzare într-o săptămînă, a mărfii provenită de la fabricile A, B, C, este respectiv $\lambda_1 = 820$, $\lambda_2 = 500$ și $\lambda_3 = 1\,280$.

Prin urmare, probabilitatea căutată, aplicînd formula (III,2.1), este

$$P_{2\,600}^{(820,500,1\,280)} = \frac{2\,600!}{820!\,500!\,1\,280!} \cdot \left(\frac{15}{49}\right)^{820} \cdot \left(\frac{13}{49}\right)^{500} \cdot \left(\frac{21}{49}\right)^{1280}.$$

Valoarea acestei probabilități este foarte mică. La acest lucru ne puteam aștepta de la început, dat fiind numărul mare de cazuri posibile.

De altminteri în practică interesează probabilitatea de a obține cazuri favorabile cuprinse între anumite limite, nu probabilitatea de a obține un număr determinat de cazuri favorabile. De aceea este necesar să însumăm probabilitățile de tipul celor menționate și să găsim diferite formule de însumare.

EXEMPLUL 13. Se aruncă un zar de 12 ori. Să se calculeze probabilitatea de a obține de două ori fiecare dintre cele șase fețe.

Este de observat că experiența aruncării unui zar este identică, din punct de vedere probabilistic, cu extragerea unei bile dintr-o urnă cu bile de șase culori, în care numărul de bile de aceeași culoare este același. Probabilitatea $p_i = \frac{1}{6}, (i = 1, 2, \ldots, 6)$, de a obține una din fețele zarului, este egală cu probabilitatea de a extrage o bilă de una din culorile bilelor aflate în urnă.

În exemplul nostru se fac 12 aruncări (echivalente cu 12 extracții dintr-o urnă); prin urmare, vom avea:

$m = 6$ (numărul stărilor — cazurilor egal posibile);

$n = 12$; din care, de cîte două ori se cere să apară aceeași față:

$$\lambda_1 = \lambda_2 = \lambda_3 = \lambda_4 = \lambda_5 = \lambda_6 = 2$$

avînd probabilitățile respective

$$p_1 = p_2 = p_3 = p_4 = p_5 = p_6 = \frac{1}{6}.$$

Probabilitatea căutată, aplicînd formula (III, 2.1), este

$$P_{12}^{(2,2,2,2,2,2)} = \frac{12!}{2!2!2!2!2!2!} \cdot \left(\frac{1}{6}\right)^2 \cdot \left(\frac{1}{6}\right)^2 \cdots \left(\frac{1}{6}\right)^2 = \frac{12!}{2^6} \cdot \left(\frac{1}{6}\right)^{12} = \frac{1\,925}{559\,872}.$$

EXEMPLUL 14. Doi jucători de șah de aceeași forță joacă un meci de 24 de partide. Se cere probabilitatea ca unul din jucători să cîștige meciul cu scorul de 13 la 11, cîștigînd 6 partide, pierzînd 4 și făcînd 14 remize, știind că probabilitatea ca o partidă să fie decisă este egală cu $\frac{1}{4}$, iar ca să fie remiză, este egală cu $\frac{1}{2}$.

În această problemă avem:

$m = 3$ numărul modurilor de a fi încheiată o partidă de șah: unul din jucători cîștigă, pierde sau face remiză, notîndu-se aceste rezultate, respectiv, cu 1, cu 0 sau cu $\frac{1}{2}$ puncte;

$n = 24$ numărul de partide, ceea ce echivalează cu 24 extracții dintr-o urnă cu bile de trei culori în aceeași proporție;

$\lambda_1 = 14$ numărul partidelor încheiate cu un rezultat de remiză, fiecare rezultat avînd probabilitatea de $\frac{1}{2}$;

$\lambda_2 = 6$ numărul de partide cîștigate de jucătorul care a cîștigat și meciul cu scorul de 13 la 11, fiecare partidă cîștigată avînd probabilitatea $p_2 = \frac{1}{4}$;

$\lambda_3 = 4$ numărul partidelor pierdute de același jucător (cîștigate de partenerul său), pentru ca scorul meciului să fie cel arătat în problemă, fiecare partidă pierdută avînd, de asemenea, probabilitatea $p_3 = \frac{1}{4}$.

Este evident că

$$\lambda_1 + \lambda_2 + \lambda_3 = 24$$

și

$$p_1 + p_2 + p_3 = \frac{1}{2} + \frac{1}{4} + \frac{1}{4} = \frac{2}{4} + \frac{1}{4} + \frac{1}{4} = \frac{4}{4} = 1.$$

Aplicînd formula (III, 2.1), probabilitatea căutată este

$$P_{24}^{(14,6,4)} = \frac{24!}{14!6!4!} \cdot \left(\frac{1}{2}\right)^{14} \cdot \left(\frac{1}{4}\right)^{6} \cdot \left(\frac{1}{4}\right)^{4}.$$

Aceasta este probabilitatea căutată pentru cazul cînd specificăm jucătorul care trebuie să cîștige. În cazul contrar, probabilitatea este egală cu dublul valorii de mai sus.

3. Schema bilei nerepetate

Se consideră o urnă conținînd A bile albe și B bile negre. Din această urnă se fac n extracții succesive, fără să se pună bilele înapoi în urnă. Să se calculeze probabilitatea ca din cele n bile extrase, α să fie albe și β negre ($\alpha + \beta = n$).

Vom vedea mai departe (vezi Cap. XX) că această schemă probabilistică prezintă interes și pentru statistica economică ori de cîte ori este rațional și posibil ca în studiul fenomenelor social-economice să se aplice metoda selecției — o metodă a statisticii matematice.

Pentru a determina probabilitatea căutată se aplică formula:

$$P = \frac{C_A^{\alpha} \cdot C_B^{\beta}}{C_{A+B}^{n}} \qquad \text{(III, 3.1)}$$

în care avem:

— La numitor, numărul combinărilor ce se pot face cu numărul total al bilelor din urnă, în raport cu numărul $\alpha + \beta = n$ al bilelor extrase;

— La numărător, numărul de succesiuni favorabile de cîte α bile albe și β bile negre ce se pot verifica (este evident că cu cele A bile albe se pot forma C_A^α grupuri de cîte α bile albe; fiecărei din aceste grupe îi putem asocia C_B^β bile negre, formate din cele B bile negre cîte sînt în urnă, luate cîte β, deci numărul succesiunilor favorabile este $C_A^\alpha \cdot C_B^\beta$).

În cazul mai general, cînd avem în urnă N bile de k culori și anume N_1 bile de culoarea întîi, N_2 bile de culoarea a doua etc. ($N_1 + N_2 + \ldots + N_k = N$), probabilitatea ca în n extracții (fără să se pună bilele înapoi în urnă) să se obțină $n_1, n_2, \ldots, n_k$ bile de culorile respective, este dată de formula:

$$P = \frac{C_{N_1}^{n_1} \cdot C_{N_2}^{n_2} \ldots C_{N_k}^{n_k}}{C_{N_1+N_2+\ldots+N_k}^{n_1+n_2+\ldots+n_k}}. \qquad \text{(III, 3.2)}$$

EXEMPLUL 15. Într-o urnă avem 100 de bile, din care 40 de culoare albă, 45 de culoare neagră și 15 de culoare roșie. Se fac 20 de extracții succesive fără să se pună bilele înapoi în urnă. Care este probabilitatea ca din cele 20 de bile extrase, 5 să fie de culoare albă, 7 de culoare neagră și 8 de culoare roșie?

În acest exemplu avem:

$N = 100$;
$N_1 = 40$; $N_2 = 45$; $N_3 = 15$; $(N_1 + N_2 + N_3 = N = 100)$;
$n = 20$;
$n_1 = 5$; $n_2 = 7$; $n_3 = 8$; $(n_1 + n_2 + n_3 = n = 20)$.

Aplicînd formula (III, 3.2), aflăm că probabilitatea căutată este

$$P = \frac{C_{40}^5 \cdot C_{45}^7 \cdot C_{15}^8}{C_{40+45+15}^{5+7+8}} = \frac{C_{40}^5 \cdot C_{45}^7 \cdot C_{15}^8}{C_{100}^{20}}.$$

4. Schema lui Poisson *)

Se consideră n urne $U_1, U_2, \ldots, U_n$, care conțin bile albe și negre în proporții date. Din fiecare urnă se extrage o bilă la întîmplare. Se cere probabilitatea de a extrage k bile albe și n-k bile negre.

Notăm cu:

$p_1, p_2, \ldots, p_n$, probabilitățile de a fi scoasă o bilă albă din urnele $U_1, U_2, \ldots, U_n$ și

$q_1, q_2, \ldots, q_n$ probabilitățile de a fi scoasă o bilă neagră din aceleași urne. Evident,

$$q_i = 1 - p_i, \qquad (i = 1, 2, \ldots, n)$$

*) S. D. Poisson (1781—1840), matematician și fizician francez, are numeroase lucrări în domeniul mecanicii cerești. De o valoare deosebită sînt contribuțiile sale în domeniul matematicilor pure. El a aprofundat studiul integralelor definite, al teoriei ecuațiilor diferențiale cu derivate parțiale și al calculului probabilităților.

Problema care se pune este aceea de a se calcula probabilitatea ca în cele n experiențe să obținem de k ori bila albă și de $n-k$ ori bila neagră.

Notăm cu $i_1, i_2, \ldots, i_k$ o combinație de cîte k numere din șirul $1, 2, 3, \ldots, n$ și cu $i_{k+1}, i_{k+2}, \ldots i_n$ o combinație de celelalte $n-k$ numere; aplicînd regulile de adunare și înmulțire a probabilităților, obținem

$$P = \Sigma\, p_{i_1} p_{i_2} \cdots p_{i_k} q_{i_{k+1}} q_{i_{k+2}} \cdots q_{i_n}. \qquad \text{(III, 4.1)}$$

Toți termenii acestei sume sînt diferitele produse parțiale ale dezvoltării

$$(p_1 + q_1) \cdot (p_2 + q_2) \cdots (p_n + q_n) = 1$$

și anume sînt acei termeni care conțin k factori p și $n-k$ factori q.

Prin urmare, dacă vom considera produsul

$$(p_1 t + q_1)(p_2 t + q_2) \cdots (p_n t + q_n),$$

probabilitatea căutată P este tocmai coeficientul lui t^k din dezvoltarea acestui produs.

EXEMPLUL 16. Avem trei urne, U_1, U_2, U_3, cu următoarea structură:

U_1 are 2 bile albe și 3 negre,
U_2 are 3 bile albe și 2 negre,
U_3 are 2 bile albe și 2 negre.

Se cere probabilitatea P ca din trei extracții (din fiecare urnă cîte o extracție) să obținem o bilă albă (deci două negre).

După cum am văzut, această probabilitate este coeficientul lui t din produsul

$$\left(\frac{2}{5}t + \frac{3}{5}\right)\left(\frac{3}{5}t + \frac{2}{5}\right)\left(\frac{1}{2}t + \frac{1}{2}\right).$$

Acest coeficient fiind egal cu

$$\frac{4}{50} + \frac{9}{50} + \frac{6}{50}$$

rezultă că

$$P = \frac{19}{50}.$$

Dacă se cere, de exemplu, probabilitatea ca să nu iasă nici o bilă albă (să iasă, deci, trei bile negre), vom calcula termenul liber al produsului, care are valoarea $\frac{6}{50}$.

5. Inegalitatea lui Boole

În capitolul II am dat regulile pe care le putem folosi pentru a determina probabilitatea ca evenimentele $E_1, E_2, \ldots, E_n$ dintr-un cîmp de evenimente independente sau dependente, să se verifice împreună.

În cazul cînd nu avem indicații asupra naturii evenimentelor E_i, probabilitatea căutată nu se poate determina printr-o formulă generală. Este însă, întotdeauna, posibil să se determine o limită inferioară a probabilității $P_{123\ldots n}$.

Pentru calculul acestei limite putem folosi metoda inducției matematice.

Considerăm evenimentele E_1 și E_2 și notăm cu:

p_1, probabilitatea ca evenimentul E_1 să se verifice singur sau împreună cu E_2;

p_2, probabilitatea ca evenimentul E_2 să se verifice singur sau împreună cu E_1;

p_{12}, probabilitatea ca E_1 și E_2 să se verifice împreună.

Aplicînd formula (II,2.4), avem

$$P = p_1 + p_2 - p_{12}$$

în care presupunem p_1, p_2 și P cunoscute, iar p_{12} necunoscută.

Deoarece avem dubla inegalitate $0 \leqslant P \leqslant 1$, putem scrie

$$p_{12} \geqslant p_1 + p_2 - 1.$$

Considerăm acum un eveniment mai mult, E_3, și notînd cu p_{123} probabilitatea evenimentului ca E_1, E_2 și E_3 să se verifice împreună, vom avea, în mod analog:

$$p_{123} \geqslant p_3 + p_{12} - 1$$

de unde, prin adunare parte cu parte, obținem

$$p_{123} \geqslant p_1 + p_2 + p_3 - 2$$

și, în general, vom avea

$$p_{123\ldots n} \geqslant p_1 + p_2 + \ldots + p_n - (n-1). \qquad \text{(III, 5.1)}$$

Această inegalitate se numește inegalitatea lui Boole.

Trebuie să precizăm aici că formula (II, 2.4) de la care am plecat pentru a ajunge la determinarea limitei inferioare a probabilității $p_{12\ldots n}$, poate fi scrisă și sub o altă formă, ce se încadrează în teoria generală a evenimentelor.

Astfel:

$$P(E_1 \cup E_2) = P(E_1) + P(E_2) - P(E_1 \cap E_2).$$

EXEMPLUL 17. Pentru ca o dare de seamă statistică să corespundă nevoilor de informare operativă și științifică, aceasta trebuie să sosească în termen la Direcția Centrală de Statistică (D.C.S.) și să cuprindă în același timp date de bună calitate și în volum complet. Dintr-un studiu statistic întocmit asupra 1 000 de dări de seamă s-a constatat că: 98%

dintre ele îndeplinesc condiția de a fi sosit la termenul fixat ; 96 %, condiția de a conține date de bună calitate și 89 %, condiția de a avea date în volum complet. Se cere probabilitatea ca o dare de seamă luată la întîmplare să îndeplinească toate cele trei condiții.

Pentru rezolvarea problemei se poate folosi inegalitatea lui Boole (III, 5.1).

Notînd cu a, b, c cele trei condiții, probabilitatea ca o dare de seamă să întrunească toate condițiile este

$$P_{a,\,b,\,c} \geqslant 0{,}98 + 0{,}96 + 0{,}89 - 2 = 0{,}83.$$

Valoarea obținută reprezintă, evident, limita inferioară a probabilității ; ea este cuprinsă, deci, între 0,83 și 0,98.

EXEMPLUL 18. Se consideră trei evenimente E_1, E_2, E_3 cu probabilitățile p_1, p_2, p_3 ca să se verifice fiecare. Se notează prin p_{12} probabilitatea ca să se verifice E_1 și E_2 împreună, prin p_{23}, probabilitatea ca să se verifice E_2 și E_3 împreună și prin p_{13}, probabilitatea ca să se verifice E_1 și E_3 împreună ; de asemenea se notează prin p_{123} probabilitatea ca să se verifice evenimentele E_1, E_2, E_3 împreună. Să se determine probabilitatea ca să se verifice cel puțin unul din evenimentele E_1, E_2, E_3.

Folosind procedeul din exemplul 7 se ajunge la concluzia că probabilitatea căutată este

$$P = p_1 + p_2 + p_3 - p_{12} - p_{23} - p_{13} + p_{123},$$

care este echivalentă cu

$$P(E_1 \cup E_2 \cup E_3) = P(E_1) + P(E_2) + P(E_3) - P(E_1 \cap E_2) - P(E_2 \cap E_3) - P(E_1 \cap E_3) + P(E_1 \cap E_2 \cap E_3).$$

Dacă notăm $p_i = 1 - q_i$, inegalitatea (III, 5.1) se poate scrie sub forma

$$p_1 p_2 \dots p_n \geqslant 1 - \sum_{i=1}^{n} q_i.$$

Și, mai departe, dacă admitem că Σq_i este o serie convergentă, inegalitatea lui Boole o putem extinde la o succesiune nelimitată de evenimente, în care caz obținem

$$p = \lim_{n \to \infty} p_{12\dots n} \geqslant 1 - \sum_{i=1}^{\infty} q_i. \qquad \text{(III, 5.2)}$$

Capitolul IV

Valori medii

1. Variabilă aleatoare. Lege de probabilitate. Valoare medie (Speranță matematică)

Să presupunem că avem o mărime X, care, variind la întîmplare, poate lua, cu probabilități cunoscute, una din valorile $x_1, x_2, \ldots, x_n$, după cum se verifică respectiv unul din evenimentele parțiale $E_1, E_2, \ldots, E_n$, pe care le presupunem incompatibile. Evenimentele E_i verifică relația

$$E_1 \cup E_2 \cup \ldots \cup E_n = E$$

în care prin E am notat, ca și pînă acum, evenimentul sigur.

Mărimea X, care variază la întîmplare, este o *variabilă aleatoare* (*întîmplătoare* sau *statistică*).

Dacă notăm cu $p_1, p_2, \ldots, p_n$ probabilitățile respective ale evenimentelor $E_1, E_2, \ldots, E_n$, vom avea, evident, în baza regulei de adunare a probabilităților,

$$p_1 + p_2 + \ldots + p_n = 1.$$

Mulțimea valorilor $x_1, x_2, \ldots, x_n$ pe care le poate lua o variabilă X, cu probabilitățile respective $p_1, p_2, \ldots, p_n$, se numește *distribuția variabilei* X, care, schematic, se poate nota

$$X\begin{bmatrix} x_1, & x_2, & \ldots, & x_n \\ p_1, & p_2, & \ldots, & p_n \end{bmatrix}.$$

Dacă, de exemplu, considerăm experiența aruncării unui zar și notăm cu X fața care apare, distribuția variabilei X este

$$X\begin{bmatrix} 1, & 2, & 3, & 4, & 5, & 6 \\ \frac{1}{6}, & \frac{1}{6}, & \frac{1}{6}, & \frac{1}{6}, & \frac{1}{6}, & \frac{1}{6} \end{bmatrix}.$$

De asemenea, putem considera ca o variabilă aleatoare numărul de bile albe ieşite dintr-o urnă Bernoulli în n extracţii succesive. Într-adevăr, în n extracţii poate să nu ne apară nici o bilă albă (deci, apar toate negre); poate să ne apară însă şi $1, 2, \ldots, n$ bile albe. Prin urmare variabila noastră aleatoare (numărul de bile albe ieşite) poate lua valorile de la 0 la n.

Probabilităţile corespunzătoare valorilor $0, 1, 2, \ldots, n$, în baza formulei (III, 1.2) pentru urna Bernoulli cu două stări, vor fi:

$C_n^0 p^0 q^{n-0} = q^n$, pentru valoarea 0 a variabilei (nici o bilă albă);
$C_n^1 p^1 q^{n-1}$ „ „ 1 „ „ (o bilă albă);
$C_n^2 p^2 q^{n-2}$ „ „ 2 „ „ (2 bile albe);
. . . . . .
$C_n^n p^n q^{n-n} = p^n$ „ „ n „ „ (n bile albe).

Variabila aleatoare considerată are deci următoarea distribuţie:

$$X\begin{pmatrix} 0 & 1 & \ldots & \alpha & \ldots & n \\ q^n & C_n^1 p\, q^{n-1} & \ldots & C_n^\alpha p^\alpha q^{n-\alpha} & \ldots & p^n \end{pmatrix}.$$

VALOARE MEDIE. Valoarea medie $M(X)$ a unei variabile aleatoare, care poate lua un număr finit de valori

$$X\begin{pmatrix} x_1, x_2, \ldots, x_n \\ p_1, p_2, \ldots, p_n \end{pmatrix}$$

este, prin definiţie,

$$M(X)^*) = p_1 x_1 + p_2 x_2 + \ldots + p_n x_n. \qquad \text{(IV,1.1)}$$

Valoarea medie mai poartă numele şi de *speranţă matematică*.

Valoarea medie are mai multe proprietăţi. Astfel, dacă notăm prin x_α cea mai mică dintre valorile $x_1, x_2, \ldots, x_n$ şi prin x_β cea mai mare dintre aceste valori, avem dubla inegalitate

$$x_\alpha \leqslant M(X) \leqslant x_\beta.$$

Într-adevăr, dacă în (IV, 1.1) înlocuim valorile $x_1, x_2, \ldots, x_n$ prin valoarea mai mică x_α, obţinem

$$M(X) \geqslant (p_1 + p_2 + \ldots + p_n) x_\alpha = x_\alpha.$$

În mod analog

$$M(X) \leqslant x_\beta.$$

În cursul procesului său de variaţie, variabila aleatoare X oscilează în jurul valorii ei medii. Să notăm cu ξ_i diferenţa

$$\xi_i = x_i - M(X) = x_i - m \qquad \text{(IV,1.2)}$$

*) Valoarea medie se mai notează şi prin $\bar{X}$.

în care, pentru ușurința scrierii, am notat prin m valoarea medie; această diferență de la valoarea medie m se numește *abatere*. Este evident că și abaterea poate lua mai multe valori, astfel că o putem defini ca fiind o variabilă aleatoare, egală cu diferența $X - M(X)$, adică

$$\xi = X - M(X). \qquad \text{(IV,1.3)}$$

Ca orice variabilă aleatoare, abaterea, are următoarea distribuție:

$$\xi \begin{pmatrix} \xi_1, \xi_2, \ldots, \xi_n \\ p_1, p_2, \ldots, p_n \end{pmatrix}.$$

Se poate demonstra ușor următoarea proprietate:
Valoarea medie a abaterii este egală cu zero.
Într-adevăr, în baza formulei (IV, 1.1) avem

$$M(\xi) = \sum_{i=1}^{n} p_i \xi_i = \sum_{i=1}^{n} p_i (x_i - m) = \sum_{i=1}^{n} p_i x_i - p_i m = m - m = 0.$$

EXEMPLUL 19. Se aruncă un zar. Să se calculeze valoarea medie a numărului de puncte ieșite.
Valoarea medie, în baza relației (IV,1.1), este egală cu

$$M(X) = 1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} + 3 \cdot \frac{1}{6} + 4 \cdot \frac{1}{6} + 5 \cdot \frac{1}{6} + 6 \cdot \frac{1}{6} = 3{,}5.$$

Calculînd abaterile $x_i - M(X)$ avem

$$1 - 3{,}5 = -2{,}5,$$
$$2 - 3{,}5 = -1{,}5,$$
$$3 - 3{,}5 = -0{,}5,$$
$$4 - 3{,}5 = +0{,}5,$$
$$5 - 3{,}5 = +1{,}5,$$
$$6 - 3{,}5 = +2{,}5.$$

Distribuția abaterii este

$$\xi \begin{bmatrix} -2{,}5, & -1{,}5, & -0{,}5, & 0{,}5, & 1{,}5, & 2{,}5 \\ \frac{1}{6}, & \frac{1}{6}, & \frac{1}{6}, & \frac{1}{6}, & \frac{1}{6}, & \frac{1}{6} \end{bmatrix}.$$

Se verifică imediat, aplicînd formula (IV,1.1), că valoarea medie a abaterii este egală cu zero.

Scanned with OKEN Scanner

2. Valoarea medie a sumei şi a produsului a două variabile aleatoare

Cînd considerăm mai multe variabile aleatoare împreună, sîntem în prezenţa unui sistem de variabile aleatoare.

Să ne limităm, pentru uşurinţa înţelegerii, la două variabile X şi Y :

$$X\begin{pmatrix} x_1, x_2, \ldots, x_m \\ p_1, p_2, \ldots, p_m \end{pmatrix}, \qquad Y\begin{pmatrix} y_1, y_2, \ldots, y_n \\ q_1, q_2, \ldots, q_n \end{pmatrix}$$

Perechea (X, Y) are următoarea distribuţie :

$$X, Y\begin{bmatrix} (x_1, y_1), (x_1, y_2), \ldots (x_1, y_n), (x_2, y_1), (x_2, y_2), \ldots (x_m y_n) \\ p_{11}, \quad p_{12}, \quad \ldots \quad p_{1n}, \quad p_{21}, \quad p_{22}, \quad \ldots \quad (p_{mn}) \end{bmatrix}, \quad \text{(IV,2.1)}$$

unde p_{ij} reprezintă probabilitatea ca să se obţină perechea (x_i, y_j).

Este evident că dacă cele două variabile sînt independente, în sensul în care am definit această noţiune, atunci

$$p_{ij} = p_i \cdot q_j ,$$

în baza regulei de înmulţire a probabilităţilor.

Variabilelor aleatoare care formează un sistem, li se pot aplica diferite operaţii ca de exemplu, înmulţirea cu o constantă K, adunarea, înmulţirea etc.

Dacă considerăm de exemplu expresia $X + Y$ — suma celor două variabile aleatoare de mai sus — aceasta este de asemenea o variabilă aleatoare care poate lua diferite valori ca $x_1 + y_1$, $x_1 + y_2$, $x_2 + y_1$, $x_2 + y_2$, ... deci în general valori $x_i + y_j$, cu probabilităţile respective $p_{11}, p_{12}, p_{21}, p_{22}, \ldots, p_{ij}$.

Este clar că problema se pune de a calcula valori medii şi pentru *sisteme de variabile aleatoare.*

a) *Valoarea medie a sumei de două variabile.* Notînd cu $i = 1, 2, \ldots, h$, indicii valorilor x_i, luate de X, şi cu $j = 1, 2, \ldots, k$ indicii valorilor y_j, luate de Y, valoarea medie pentru suma $X + Y$, o putem defini în baza relaţiei (IV,1,1), astfel :

$$M(X + Y) = p_{11}(x_1 + y_1) + p_{12}(x_1 + y_2) + \ldots + p_{h1}(x_h + y_1) +$$
$$+ p_{h2}(x_h + y_2) + \ldots + p_{hk}(x_h + y_k).$$

Termenii din partea a doua a egalităţii de mai sus îi putem grupa convenabil, obţinînd

$$M(X + Y) = x_1(p_{11} + p_{12} + \ldots + p_{1k}) +$$
$$+ \ldots + x_h(p_{h1} + p_{h2} + \ldots + p_{hk}) + \ldots + y_k(p_{1k} + \ldots + p_{hk}),$$

de unde, ţinînd seama de relaţiile (IV,2.1), deducem

$$M(X + Y) = p_1 x_1 + \ldots + p_h x_h + q_1 y_1 + \ldots + q_k y_k$$

și cum, în baza formulei (IV,1.1),

$$p_1 x_1 + \ldots + p_h x_h = M(X) \quad \text{și} \quad q_1 y_1 + \ldots + q_k y_k = M(Y),$$

avem

$$M(X+Y) = M(X) + M(Y). \qquad \text{(IV,2.2)}$$

Putem deci formula următoarea teoremă:

Valoarea medie a unei sume de două variabile aleatoare dintr-un sistem, este egală cu suma valorilor medii ale celor două variabile aleatoare.

b) *Valoarea medie pentru un număr finit de variabile.* În general, pentru n variabile aleatoare $X_1, X_2, \ldots, X_n$ — urmînd același raționament ca în cazul sumei a două variabile — media sumei sistemului de variabile independente este:

$$M(X_1 + X_2 + \ldots + X_n) = M(X_1) + M(X_2) + \ldots + M(X_n). \qquad \text{(IV, 2.3)}$$

Așadar în cazul considerat putem formula următoarea teoremă:

Valoarea medie a sumei unui număr finit de variabile aleatoare este egală cu suma valorilor medii ale fiecărei variabile.

c) *Valoarea medie a unei variabile din sistem înmulțită cu o constantă K,* este egală cu:

$$M(KX) = KM(X). \qquad \text{(IV,2.4)}$$

3. Valoarea medie a produsului variabilelor independente dintr-un sistem

Dacă notăm cu Z produsul celor două variabile aleatoare X și Y considerate, adică

$$Z = X \cdot Y,$$

obținem o altă variabilă aleatoare pentru care de asemenea se pune problema de a calcula media.

Vom avea, în baza relațiilor privitoare la valoarea medie a unei variabile aleatoare:

$$M(Z) = M(X \cdot Y) = p_{11} x_1 y_1 + p_{12} x_1 y_2 + \ldots + p_{1k} x_1 y_k + \ldots +$$
$$+ p_{21} x_2 y_1 + p_{22} x_2 y_2 + \ldots + p_{h1} x_h y_1 + p_{h2} x_h y_2 + \ldots +$$
$$+ p_{hk} x_h y_k = p_1 x_1 q_1 y_1 + p_1 x_1 q_2 y_2 + \ldots + p_1 x_1 q_k y_k + \ldots +$$
$$+ p_2 x_2 q_1 y_1 + p_2 x_2 q_2 y_2 + \ldots + p_h x_h q_1 y_1 + p_h x_h q_2 y_2 + \ldots +$$
$$+ p_h x_h q_k y_k = (p_1 x_1 + p_2 x_2 + \ldots + p_h x_h)(q_1 y_1 + q_2 y_2 + \ldots + q_k y_k) =$$
$$= M(X)\, M(Y). \qquad \text{(IV,3.1)}$$

În general, pentru n variabile independente $X_1, X_2, \ldots, X_n$ avem

$$M(X_1 X_2 \ldots X_n) = M(X_1)\, M(X_2) \ldots M(X_n). \qquad \text{(IV,3.2)}$$

Scanned with OKEN Scanner

Prin urmare putem enunța următoarea teoremă:

Valoarea medie a unui produs de variabile independente este egală cu produsul valorilor medii ale variabilelor considerate.

Trebuie să subliniem că relațiile (IV, 2.3) și (IV,3.2) au o mare importanță, deoarece prin aplicarea lor se poate determina mai simplu valoarea medie în diferite cazuri în care calculul direct este mai greu de aplicat.

4. Valoarea medie a numărului de bile albe obținut în n extracții succesive dintr-o urnă Bernoulli

EXEMPLUL 20. Se consideră o urnă Bernoulli cu două stări. Cunoaștem deci probabilitatea p ca să scoatem o bilă albă și probabilitatea q de a scoate o bilă neagră ($p + q = 1$). Se fac n extracții succesive și se cere să se determine valoarea medie a numărului de bile albe ieșite.

Însemnăm prin X_1 numărul de bile albe pe care-l putem obține în prima extracție. Este evident că dintr-o singură extracție nu putem scoate decît o bilă albă sau nici una. Variabila X_1 nu poate lua decît valorile 1 sau 0 cu probabilitățile respective p sau q.

De asemenea, să însemnăm prin X_2 numărul de bile albe obținute la a doua extracție. La fel, X_2 nu poate lua decît valorile 1 sau 0 după cum am obținut în această extracție o bilă albă sau una neagră.

Continuînd astfel, avem un șir de variabile aleatoare:

$$X_1, X_2, X_3, \ldots, X_n$$

care reprezintă numărul de bile albe obținut în extracțiile de rangul i, ($i = 1, 2, 3, \ldots, n$).

Numărul de bile albe obținut în cele n extracții se capătă prin adunarea numărului de bile obținut în fiecare extracție. El rezultă astfel prin însumarea variabilelor $X_1, X_2, \ldots, X_n$, adică $X_1 + X_2 + \ldots + X_n$.

Însemnînd însă prin $M(X_1 + X_2 + \ldots + X_n)$ valoarea medie a sumei $X_1 + X_2 + \ldots + X_n$, avem, în baza relației (IV, 2.3), relația

$$M(X_1 + X_2 + \ldots + X_n) = M(X_1) + M(X_2) + \ldots + M(X_n). \quad \text{(IV,4.1)}$$

În această relație vom avea întîi pe $M(X_1)$.

În baza definiției date valorii medii (IV,1.1), $M(X_1)$ este dată de relația

$$M(X_1) = p \cdot 1 + q \cdot 0 = p$$

deoarece, după cum am văzut mai sus, X_1 poate lua numai valorile 1 și 0, cu probabilitățile respective p și q. În mod analog avem

$$M(X_2) = p\,;\ M(X_3) = p\,;\ \ldots\ M(X_n) = p.$$

Înlocuind aceste valori în relația (IV, 4.1), obținem

$$M(X_1 + X_2 + \ldots + X_n) = p + p + \ldots + p = n \cdot p. \quad \text{(IV, 4.2)}$$

Prin urmare putem formula următoarea regulă:

Valoarea medie a numărului de bile albe obținute din n extracții succesive, făcute în condițiile urnei lui Bernoulli, este egală cu np, adică

$$M(X) = \sum_{i=1}^{n} M(X_i) = np. \qquad \text{(IV,4.3)}$$

La același rezultat am fi putut ajunge considerînd distribuția variabilei aleatoare X, a numărului de bile albe ieșite în n extracții.

Însemnînd cu α numărul de bile albe ieșite în n extracții, știm că probabilitatea acestui eveniment, în baza formulei (III,1.2), este egală cu

$$P_{n,\alpha} = C_n^{\alpha} p^{\alpha} q^{n-\alpha}.$$

Distribuția corespunzătoare este

$$X\begin{bmatrix} 0 & 1 & 2 & \dots & n \\ C_n^0 p^0 q^{n-0} & C_n^1 p^1 q^{n-1} & C_n^2 p^2 q^{n-2} & \dots & C_n^n p^n q^{n-n} \end{bmatrix}$$

sau în formă concentrată

$$X\left[i, C_n^i p^i q^{n-i} \right], \qquad (i = 0, 1, 2, 3, \dots, n).$$

Aplicînd formula (IV, 1.1) pentru medie, vom avea

$$M(X) = \sum_{i=1}^{n} i C_n^i p^i q^{n-i}. \qquad \text{(IV,4.4)}$$

Pentru a calcula suma din partea dreaptă a relației (IV,4.4), considerăm identitatea

$$\sum_{i=0}^{n} C_n^i p^i t^i q^{n-i} = (pt + q)^n,$$

în care t este un parametru oarecare.

Derivînd această relație în raport cu t, deducem

$$\sum_{i=0}^{n} i C_n^i p^i t^{i-1} q^{n-1} = np(pt + q)^{n-1}.$$

Dacă facem $t = 1$ și ținem seama că $p + q = 1$, rezultă

$$\sum_{i=0}^{n} i C_n^i p^i t^i q^{n-1} = np,$$

adică aceeași valoare pe care am obținut-o mai sus.

În legătură cu valoarea medie definită prin relația (IV,4.3) putem face o observație cu caracter general și anume: ori de cîte ori considerăm un eveniment a cărui probabilitate de verificare în fiecare probă este p, rezultatul obținut se păstrează fără nici o modificare. Putem astfel afirma că valoarea medie a numărului de cîte ori s-a verificat un eveniment a cărui probabilitate de verificare este p, cînd se repetă de n ori, este egală cu np.

EXEMPLUL 21. Se aruncă o monedă de șapte ori. Să se calculeze valoarea medie a numărului în care am obținut fața cu banul.

Sîntem în cazul urnei lui Bernoulli cu două stări, cu probabilităţile $p=\frac{1}{2}$ şi $q=\frac{1}{2}$. Evident, $p+q=1$. Din această urnă se fac $n=7$ extracţii.

Variabila aleatoare X (numărul de cazuri de apariţie a feţei cu banul), în cele şapte aruncări cu banul poate lua valori de la 0 la 7 inclusiv. Vom avea deci distribuţia :

$$X\begin{bmatrix} 0, & 1, & 2, & 3, & 4, & 5, & 6, & 7 \\ q^7, & C_7^1 p^1 q^6, & C_7^2 p^2 q^5, & C_7^3 p^3 q^4, & C_7^4 p^4 q^3, & C_7^5 p^5 q^2, & C_7^6 p^6 q^1, & C_7^7 p^7 q^0 \end{bmatrix}.$$

în care făcînd înlocuirile corespunzătoare se obţine

$$\begin{bmatrix} 0, & 1, & 2, \\ \left(\frac{1}{2}\right)^7, & 7\cdot\frac{1}{2}\cdot\left(\frac{1}{2}\right)^6, & 21\cdot\left(\frac{1}{2}\right)^2\left(\frac{1}{2}\right)^5, & 35\cdot\Big(\end{bmatrix}$$

sau în formă şi mai simplificată, făcînd şi calculul probabilităţilor,

$$X\begin{bmatrix} 0, & 1, & 2, & 3, & 4, & 5, & 6, & 7 \\ \frac{1}{128}, & \frac{7}{128}, & \frac{21}{128}, & \frac{35}{128}, & \frac{35}{128}, & \frac{21}{128}, & \frac{7}{128}, & \frac{1}{128} \end{bmatrix}.$$

Pentru a afla valoarea medie a acestei distribuţii putem aplica formula (IV,1.1), în care caz obţinem

$$M(X)=1\cdot\frac{7}{128}+2\cdot\frac{21}{128}+3\cdot\frac{35}{128}+4\cdot\frac{35}{128}+5\cdot\frac{21}{128}+6\cdot\frac{7}{128}+7\cdot\frac{1}{128}=$$

$$\frac{448}{128}=3{,}5.$$

Cunoscînd, însă, valoarea medie a distribuţiei binomiale ca fiind egală cu np, la acelaşi rezultat am fi putut ajunge mult mai simplu, aplicînd formula (IV, 4.3). Deci

$$M(X)=\sum_{i=1}^{n} M(X_i)=np=7\cdot\frac{1}{2}=\frac{7}{2}=3{,}5.$$

5. Momente. Media armonică. Media geometrică

Am precizat mai înainte că asupra variabilelor aleatoare se pot face anumite operaţii, obţinîndu-se alte variabile aleatoare care au distribuţiile lor proprii cu valorile medii corespunzătoare. O variabilă aleatoare poate fi ridicată, deci, şi la o putere de ordinul k, în care caz se pune problema de a-i calcula valoarea medie.

Scanned with OKEN Scanner

Sîntem în cazul urnei lui Bernoulli cu două stări, cu probabilităţile $p=\frac{1}{2}$ şi $q=\frac{1}{2}$. Evident, $p+q=1$. Din această urnă se fac $n=7$ extracţii.

Variabila aleatoare X (numărul de cazuri de apariţie a feţei cu banul), în cele şapte aruncări cu banul poate lua valori de la 0 la 7 inclusiv. Vom avea deci distribuţia :

$$X\begin{bmatrix} 0, & 1, & 2, & 3, & 4, & 5, & 6, & 7 \\ q^7, & C_7^1 p^1 q^6, & C_7^2 p^2 q^5, & C_7^3 p^3 q^4, & C_7^4 p^4 q^3, & C_7^5 p^5 q^2, & C_7^6 p^6 q^1, & C_7^7 p^7 q^0 \end{bmatrix}.$$

în care făcînd înlocuirile corespunzătoare se obţine

$$\begin{bmatrix} 0, & 1, & 2, & 3, & 4, & 5, & 6, & 7 \\ \left(\frac{1}{2}\right)^7, & 7\cdot\frac{1}{2}\cdot\left(\frac{1}{2}\right)^6, & 21\cdot\left(\frac{1}{2}\right)^2\left(\frac{1}{2}\right)^5, & 35\cdot\left(\frac{1}{2}\right)^3\left(\frac{1}{2}\right)^4, & 35\cdot\left(\frac{1}{2}\right)^4\left(\frac{1}{2}\right)^3, & 21\cdot\left(\frac{1}{2}\right)^6\left(\frac{1}{2}\right)^2, & 7\cdot\left(\frac{1}{2}\right)^6\frac{1}{2}, & \left(\frac{1}{2}\right)^7 \end{bmatrix}$$

sau în formă şi mai simplificată, făcînd şi calculul probabilităţilor,

$$X\begin{bmatrix} 0, & 1, & 2, & 3, & 4, & 5, & 6, & 7 \\ \frac{1}{128}, & \frac{7}{128}, & \frac{21}{128}, & \frac{35}{128}, & \frac{35}{128}, & \frac{21}{128}, & \frac{7}{128}, & \frac{1}{128} \end{bmatrix}.$$

Pentru a afla valoarea medie a acestei distribuţii putem aplica formula (IV,1.1), în care caz obţinem

$$M(X)=1\cdot\frac{7}{128}+2\cdot\frac{21}{128}+3\cdot\frac{35}{128}+4\cdot\frac{35}{128}+5\cdot\frac{21}{128}+6\cdot\frac{7}{128}+7\cdot\frac{1}{128}=$$

$$\frac{448}{128}=3{,}5.$$

Cunoscînd, însă, valoarea medie a distribuţiei binomiale ca fiind egală cu np, la acelaşi rezultat am fi putut ajunge mult mai simplu, aplicînd formula (IV, 4.3). Deci

$$M(X)=\sum_{i=1}^{n} M(X_i)=np=7\cdot\frac{1}{2}=\frac{7}{2}=3{,}5.$$

5. Momente. Media armonică. Media geometrică

Am precizat mai înainte că asupra variabilelor aleatoare se pot face anumite operaţii, obţinîndu-se alte variabile aleatoare care au distribuţiile lor proprii cu valorile medii corespunzătoare. O variabilă aleatoare poate fi ridicată, deci, şi la o putere de ordinul k, în care caz se pune problema de a-i calcula valoarea medie.

Scanned with OKEN Scanner

Prin definiție, expresia

$$M(X^k) = p_1 x_1^k + p_2 x_2^k + \ldots + p_n x_n^k \qquad \text{(IV,5.1)}$$

este tot o valoare medie a variabilei X însă ridicată la puterea k. Unei asemenea valori medii i se spune *moment de ordinul k* al variabilei X.

Dacă se extrage radicalul de ordinul k din expresia (IV,5.1), obținem o altă expresie

$$M_k = \sqrt[k]{M(X^k)} = \sqrt[k]{p_1 x_1^k + p_2 x_2^k + \ldots + p_n k_n^k} \qquad \text{(IV,5.2)}$$

care se numește *valoarea medie de ordinul k* a aceleiași variabile.

Putem observa acum că *speranța matematică* $M(X)$ nu este decît un caz particular al expresiei (IV, 5.1), pentru $k = 1$; este vorba de momentul de ordinul întîi al variabilei aleatoare X.

Pentru $k = -1$, avem

$$M_{-1} = \left(\frac{p_1}{x_1} + \frac{p_2}{x_2} + \cdots + \frac{p_n}{x_n}\right)^{-1} \qquad \text{(IV,5.3)}$$

numită *media armonică* a valorilor x_i.

Cînd $k = 0$, momentul M_0 se prezintă sub formă nedeterminată.

Aplicînd logaritmii și regula lui L'Hôpital, se demonstrează ușor că:

$$M_0 = x_1^{p_1} \cdot x_2^{p_2} \ldots x_n^{p_n}, \qquad \text{(IV,5.4)}$$

expresie care se numește *media geometrică* a valorilor x_i, $(i = 1,2,3,\ldots,n)$.

Într-adevăr, logaritmînd expresia (IV,5.2) obținem

$$\lg M_k = \frac{1}{k} \lg (p_1 x_1^k + \ldots + p_n x_n^k)$$

iar regula lui L'Hôpital ne dă

$$\lim_{k \to 0} \lg M_k = \lim_{k \to 0} \frac{p_1 x_1^k \lg x_1 + \ldots + p_n x_n^k \lg x_n}{p_1 x_1^{-k} + \ldots + p_n x_n^k} =$$

$$= \frac{p_1 \lg x_1 + \ldots + p_n \lg x_n}{p_1 + \ldots + p_n}.$$

Deci

$$\lg M_0 = \lg x_1^{p_1} x_2^{p_2} \ldots x_n^{p_n},$$

de unde

$$M_0 = x_1^{p_1} \cdot x_2^{p_2} \ldots x_n^{p_n}.$$

Momentul de *ordinul al doilea* al unei variabile aleatoare este prin definiție:

$$M(X^2) = p_1 x_1^2 + p_2 x_2^2 + \ldots + p_n x_n^2. \qquad \text{(IV,5.5)}$$

Scanned with OKEN Scanner

Expresia $\sqrt{M(X^2)}$ se numește *valoare medie pătratică*.

Valoarea medie pătratică a variabilei X este mai mare decît valoarea medie de ordinul întîi. Se demonstrează ușor că avem relația

$$\sqrt{M(X^2)} \geqslant |M(X)|. \qquad \text{(IV, 5.6)}$$

Pentru a demonstra relația de mai sus este suficient să arătăm că

$$M(X^2) \geqslant [M(X)]^2.$$

Vom folosi cunoscuta identitate a lui Lagrange:

$$\sum a_i^2 \sum b_i^2 - (\sum a_i b_i)^2 = \sum (a_i b_j - a_j b_i)^2,$$

unde luăm $a_i = \sqrt{p_i}$, $b_i = \sqrt{p_i}\, x_i$

$$\sum p_i \sum p_i x_i^2 - (\sum p_i x_i)^2 \equiv \sum p_i p_j (x_i - x_j)^2.$$

Cum însă $\sum p_i = 1$; $\sum p_i x_i^2 = M(X^2)$; $\sum p_i x_i = M(X)$, iar suma din membrul al doilea este pozitivă, obținem inegalitatea

$$M(X^2) - [M(X)]^2 \geqslant 0$$

ceea ce, evident, înseamnă că media pătratică a variabilei X este mai mare decît momentul de ordinul întîi al aceleiași variabile. Se vede că inegalitatea de mai sus nu se poate transforma în egalitate decît dacă toate valorile x_i sînt egale între ele, ceea ce este imposibil.

6. Valoarea medie pătratică a unei sume de variabile independente dintr-un sistem

Se consideră n variabile independente $X_1, X_2, \ldots, X_n$; dacă

$$M(X_1) = M(X_2) = \ldots = M(X_n) = 0,$$

atunci vom avea următoarea relație, care constituie o cunoscută teoremă în calculul probabilităților,

$$M[(X_1 + X_2 + \ldots + X_n)^2] = M(X_1^2) + M(X_2^2) + \ldots + M(X_n^2) = \sum M(X_i^2). \qquad \text{(IV, 6.1)}$$

Teorema poate fi enunțată astfel:

Valoarea medie a pătratului unei sume de variabile independente, fiecare din ele avînd valoarea medie zero, este egală cu suma valorilor medii a pătratului fiecărei variabile în parte.

Relația (IV,6.1) se poate deduce simplu, dacă scriem:

$$M[(X_1 + X_2 + \ldots + X_n)^2] = M\left(\sum X_i^2 + 2 \sum X_i X_j\right), \qquad \text{(IV, 6.2)}$$

În partea doua a egalității se vede clar că avem media unui polinom ridicat la pătrat. Aplicînd teorema cu privire la valoarea medie a unei sume, observăm că

$$M\left(\sum X_i^2 + 2\sum X_i X_j\right) = M\left(\sum X_i^2\right) + M\left(2\sum X_i X_j\right)$$

însă

$$M\left(2\sum X_i X_j\right) = 2\,M\left(\sum X_i X_j\right) = 2\sum M\left(X_i X_j\right) = 2\sum M\left(X_i\right) M\left(X_j\right) = 0,$$

ceea ce înseamnă că avem, transformînd relația (IV,6.2),

$$M\left[(X_1 + X_2 + \ldots + X_n)^2\right] = M\left(\sum X_i^2\right)$$

sau

$$M\left[(X_1 + X_2 + \ldots + X_n)^2\right] = \sum M\left(X_i^2\right),$$

adică tocmai relația (IV,6 ·1).

7. Inegalități între valorile medii

Am arătat mai sus (IV,5.5) că valoarea medie pătratică este mai mare decît media de ordinul întîi.

Se poate demonstra că media de ordinul întîi este mai mare decît media geometrică, iar aceasta din urmă este mai mare decît media armonică, cu condiția ca valorile luate de variabila aleatoare respectivă să fie pozitive. Avem deci inegalitățile :

$$M_{-1}\left(X_i\right) \leqslant M_0\left(X_i\right) \leqslant M\left(X_i\right) \leqslant M\left(X_i^2\right),$$

Șirul acestor inegalități poate continua la infinit, pentru oricare din numerele puterilor de ordinul k, k', k'', ... care satisfac inegalitățile

$$k < k' < k'' \ldots$$

Vom avea astfel întotdeauna

$$M_k < M_{k'} < M_{k''} \ldots$$

dacă valorile $X_1, X_2, \ldots, X_n$ sînt pozitive. Am notat

$$M_k = M\left(X_i^k\right);\ M_{k'} = M\left(X_i^{k'}\right);\ M_{k''} = M\left(X_i^{k''}\right) \text{ etc.}$$

Vom demonstra că dacă $y_i > 0$, $(i = 1, 2, \ldots, n)$, avem întotdeauna

$$\left(\sum_1^n p_i y_i^r\right)^{1/r} \geqslant \sum_1^n p_i y_i, \qquad (r > 1).$$

Să însemnăm

$$\sum_1^n p_i y_i = T. \qquad (\alpha)$$

Ne propunem să găsim valorile extreme ale funcției

$$z = \left(\sum_1^n p_i y_i^r \right)^{1/r}$$

dacă este îndeplinită condiția (α).

Observăm că z este o funcție de $n-1$ variabile $y_1, y_2, \ldots, y_{n-1}$, deoarece y_n se exprimă în raport cu aceste variabile în baza relației (α).

Derivînd obținem

$$\frac{\delta z}{\delta y_i} = \frac{1}{r} p_i \left(y_i^{r-1} - y_n^{r-1} \right) \left(\sum p_i y_i^r \right)^{\frac{1}{r}-1}, \qquad (i = 1, 2, \ldots, n-1)$$

$$\frac{\delta^2 z}{\delta y_i \, \delta y_j} = \frac{1}{r} \left(\frac{1}{r} - 1 \right) \left(y_i^{r-1} - y_n^{r-1} \right) \left(y_j^{r-1} - y_n^{r-1} \right) \left(\sum p_i y_i \right)^{\frac{1}{r}-2} +$$

$$+ \frac{r-1}{r} \frac{p_i p_j}{P_n} y_n^{r-2} \left(\sum p_i y_i^r \right)^{\frac{1}{r}-1}, \qquad (i, j = 1, 2, \ldots, n-1)$$

Egalînd primele derivate cu zero, obținem

$$y_1^{r-1} = y_2^{r-1} = \ldots = y_n^{r-1}$$

egalitate care, ținînd seama de faptul că $y_i > 0$, $(i = 1, 2, \ldots, n)$, nu poate avea loc decît dacă

$$y_1 = y_2 = \ldots = y_n .$$

Ținînd seama de (α) rezultă mai departe:

$$y_i = T, \qquad (i = 1, 2, \ldots, n)$$

Pentru aceste valori z ia valoarea T.

Prin urmare, T este o valoare extremă pentru funcția z.

Pentru a deduce natura acestui extrem, observăm că în punctul $y_i = T$, $(i = 1, 2, \ldots, n)$, derivatele de ordinul al doilea iau valorile:

$$\left[\frac{\delta^2 z}{\delta y_i \, \delta y_j} \right]_T = \frac{r-1}{rT} \frac{p_i p_j}{p_n} .$$

Forma pătratică

$$\sum_{i,j=1}^{n-1} \left[\frac{\delta^2 z}{\delta y_i \, \delta y_j} \right]_T \xi_i \xi_j .$$

este pozitiv definită.

Într-adevăr, avem:

$$\sum_{i,j=1}^{n-1} \left[\frac{\delta^2 z}{\delta y_i \, \delta y_j} \right]_T \xi_i \xi_j = \frac{r-1}{r T p_n} \sum_{i,j=1}^{n-1} p_i p_j \xi_i \xi_j = \frac{r-1}{r p_n} T^{-1} \left(\sum_{i=1}^{n-1} p_i \xi_i \right)^2 \geqslant 0, \quad (r > 1).$$

Prin urmare, în punctul considerat avem un minim și deci proprietatea este demonstrată.

Dacă aplicăm substituția

$$y_i = x_i^s, \qquad (i = 1, 2, \ldots, n), \quad (s > 0)$$

din inegalitatea

$$\left(\sum_{i=1}^n p_i y_i^r \right)^{1/r} \geqslant \sum_{i=1}^n p_i y_i , \qquad (r > 1)$$

rezultă :

$$\left(\sum_{i=1}^{n} p_i\, x_i^{rs}\right)^{1/r} \geqslant \sum_{i=1}^{n} p_i\, x_i^{s},$$

$$\left(\sum p_i\, x_i^{rs}\right)^{1/rs} \geqslant \left(\sum p_i\, x_i^{s}\right)^{1/s}.$$

Cum $rs > s$, rezultă că am demonstrat că avem întotdeauna

$$\left(\sum p_i\, x_i^{p}\right)^{1/p} \geqslant \left(\sum p_i\, x_i^{q}\right)^{1/q},$$

pentru $p > q \geqslant 0$, $x_i > 0$, $(i = 1, 2, \ldots, n)$

Efectuînd substituția

$$y_i = x_i^{-s}, \qquad (i = 1, 2, \ldots, n)\ (s > 0)$$

din inegalitatea

$$\left(\sum p_i\, y_i^{r}\right)^{1/r} \geqslant \sum p_i\, y_i$$

obținem succesiv

$$\left(\sum p_i\, x_i^{-rs}\right)^{1/r} \geqslant \sum p_i\, x_i^{-s}, \qquad (r > 1,\ s > 0)$$

$$\left(\sum p_i\, x_i^{-rs}\right)^{-\frac{1}{rs}} \leqslant \left(\sum p_i\, x_i^{-s}\right)^{-\frac{1}{s}}.$$

Însă $-rs < -s$, așa că am demonstrat că avem întotdeauna

$$\left(\sum p_i\, x_i^{p}\right)^{1/p} \geqslant \left(\sum p_i\, x_i^{q}\right)^{1/q}$$

pentru $q < p \leqslant 0$, $x_i > 0$, $(i = 1, 2, \ldots, n)$.

Putem deci afirma că avem în general

$$\left(\sum p_i\, x_i^{p}\right)^{1/p} \geqslant \left(\sum p_i\, x_i^{q}\right)^{1/q}, \qquad (x_i > 0\,;\ i = 1, 2, \ldots, n)$$

pentru toate numerele reale p și q, care satisfac relația $p > q$.

8. Considerații generale asupra valorilor medii ale variabilelor aleatoare. Semnificația valorilor medii

O întrebare firească ce se poate pune este aceea de a ști la ce ne poate folosi cunoașterea valorii medii a unei variabile aleatoare. Întrebarea prezintă îndeosebi un interes practic.

Să considerăm cîteva exemple care ne vor ajuta să răspundem la o asemenea întrebare.

EXEMPLUL 22. Se organizează o olimpiadă de matematici pentru elevii din clasa XI a două școli medii A și B din aceeași regiune. Olimpiada se desfășoară în următoarele condiții :

— pentru fiecare răspuns scris care a obținut calificativul „foarte bine" se acordă 30 de puncte ;

— pentru un răspuns care a obținut calificativul „bine" se acordă 20 de puncte ;

— se acordă 10 puncte pentru un răspuns calificat „satisfăcător" și, în sfîrșit

— se acordă 1 punct pentru un răspuns calificat „slab".

Ni se cere să determinăm un indicator numeric care, prin *mărimea lui*, să ne indice în mod direct, calitatea cunoştiinţelor de matematici ale elevilor din şcolile A şi B separat şi în general ale întregii colectivităţi de elevi, cunoscîndu-se dinainte că numărul de puncte pe care-l poate obţine răspunsul la o singură întrebare dat de un elev din şcoala A are o lege de probabilitate diferită de cea a numărului de puncte pe care-l poate obţine răspunsul unui elev din şcoala B la o singură întrebare.

Sînt date, anume, următoarele legi de probabilitate ale celor două variabile aleatoare X (numărul de puncte obţinut la răspunsul unui elev din şcoala A) şi Y (numărul de puncte obţinut la un răspuns dat de un elev din şcoala B) :

$$X\begin{bmatrix} 1 & 10 & 20 & 30 \\ 0{,}1 & 0{,}1 & 0{,}4 & 0{,}4 \end{bmatrix}, \qquad \text{(IV,8.1)}$$

$$Y\begin{bmatrix} 1 & 10 & 20 & 30 \\ 0 & 0{,}2 & 0{,}5 & 0{,}3 \end{bmatrix}. \qquad \text{(IV,8.2)}$$

Examinarea nemijlocită a acestor două legi de probabilitate nu ne oferă posibilitatea să caracterizăm decît calitatea cunoştinţelor, prin unul din numerele punctelor pe care le poate obţine un elev, care — după cum se vede — poate fi or 1, or 10, or 20, or 30. Într-adevăr, un răspuns foarte bun, valorînd 30 de puncte, este mai probabil la un elev din şcoala A decît la un elev din şcoala B, dar, totodată şi un răspuns slab este de asemenea mai probabil la un elev din A decît la un elev din B ; dimpotrivă, răspunsuri bune şi satisfăcătoare sînt mai probabile la un elev din B.

Să calculăm acum — după regula cunoscută (IV,1.1) — valoarea medie a numărului, de puncte, separat pentru un elev din A şi separat pentru un elev din B.

Un calcul elementar ne arată că :

$$M(X) = 21{,}1 \text{ puncte},$$

$$M(Y) = 21{,}0 \text{ puncte}.$$

După cum se vede, un elev din A are, în medie, un număr de puncte ceva mai mare decît un elev din B. Deşi este vorba de o diferenţă nesemnificativă, totuşi aceasta înseamnă că practic, la un număr mare de răspunsuri, un elev din A va da, în general, răspunsuri mai bune decît un elev din B din aceeaşi clasă. Este evident că datele sînt strict exemplificative.

Aşadar, valoarea medie a numărului de puncte realizate ne-a dat *o măsură* comodă de comparare a calităţii cunoştinţelor de matematici ale elevilor.

Pentru a caracteriza întreaga colectivitate a elevilor concurenţi putem calcula numărul mediu de puncte ce se poate realiza, răspunzîndu-se la o întrebare, de către elevii ambelor şcoli.

Este vorba deci de a calcula valoarea medie a sumei diferitelor valori pe care le pot lua cele două variabile aleatoare X şi Y.

Scanned with OKEN Scanner

Știm că suma variabilelor aleatoare dintr-un sistem este și ea o variabilă aleatoare și că valoarea ei medie este întotdeauna egală cu suma valorilor medii ale termenilor ei (IV,2.3). Aplicînd această teoremă aflăm că :

$$M(X) + M(Y) = 21{,}1 + 21{,}0 = 42{,}1 \text{ puncte.} \qquad \text{(IV,8.3)}$$

Este evident că făcînd un număr mic de experiențe practice, noi putem obține probabilități statistice și valori medii statistice care se vor abate de la legea de probabilitate teoretică și respectiv de la valoarea medie teoretică, însă, după cum vom vedea mai departe, în cazul unui număr mare de probe, mediile statistice vor oscila cu abateri nesemnificative în jurul valorii teoretice, iar legitatea statistică se va apropia de legea de probabilitate.

Semnificația valorii medii a unei variabile aleatoare devine astfel clară. Din punct de vedere practic, valoarea medie caracterizează foarte bine variabila aleatoare, atunci cînd este vorba de o operație de *masă*, sau *repetată* de multe ori.

EXEMPLUL 23. Se cere să se găsească valoarea medie a ariei unui teren dreptunghiular, știind că valoarea medie a măsurătorilor unei laturi este a și că valoarea medie a măsurătorilor celeilalte laturi este b.

Măsurarea exactă a unui segment nu este cu putință. Cu toate că măsurarea unui segment este făcută în cele mai bune condiții, de fiecare dată se obține un alt rezultat al măsurătorii, datorită impreciziei instrumentelor, imperfecțiunii simțurilor omenești etc.

Diversele rezultate ale măsurătorii se distribuie în jurul unei valori medii, după o lege de probabilitate anumită. Dacă notăm prin x valoarea obținută prin măsurare a unei laturi a dreptunghiului, x este o variabilă aleatoare, a cărei valoare medie, în baza datelor problemei, o cunoaștem :

$$M(x) = a.$$

La fel, notînd prin y cealaltă latură a dreptunghiului avem :

$$M(y) = b.$$

Aria dreptunghiului este xy.

x și y fiind variabile aleatoare, înseamnă că și aria xy este o variabilă aleatoare.

Dacă x și y sînt variabile independente, atunci

$$M(x\ y) = M(x)\ M(y) = ab.$$

Valoarea medie a ariei este ab.

Esențial în obținerea acestui rezultat este faptul că x și y sînt independente între ele. Se vede mai clar acest lucru, presupunînd că în loc de dreptunghi avem un pătrat. Fie

$$M(x) = a,$$

valoarea medie a laturii pătratului. Aria pătratului este x^2 şi avem

$$M(x^2) \neq M(x) \cdot M(x).$$

Procedeul aplicat în cazul dreptunghiului nu mai este aplicabil. Mai ştim că

$$M(x^2) \geqslant [M(x)]^2.$$

Deci valoarea medie a ariei pătratului este întotdeauna mai mare decît pătratul valorii medii a laturii lui.

Şi în acest exemplu funcţia mediei rămîne aceeaşi : de a evalua printr-un singur număr, nivelul mediu real al variabilei aleatoare care ia o mulţime de valori.

În general media este o valoare *abstractă*, dedusă din valorile date ale variabilei. Aceasta nu înseamnă, însă, că ea este o valoare fictivă; dimpotrivă, caracterul ei este prin excelenţă *real* şi în acelaşi timp *obiectiv* şi *necesar*, după cum reală, obiectivă şi necesară este şi legea de probabilitate a variabilei aleatoare pe care o caracterizează.

Trebuie să precizăm aici că regula dedusă la (IV,3.2), în sensul că

$$M(x) \cdot M(y) = M(x \cdot y),$$

sau în altă notaţie

$$\bar{x} \cdot \bar{y} = \overline{x \cdot y},$$

nu este generală. Cunoaşterea valorilor medii $\bar{x}$ şi $\bar{y}$ nu determină în general, o valoare a lui $\overline{xy}$. Aceasta înseamnă că pentru unele şi aceleaşi $\bar{x}$ şi $\bar{y}$, putem obţine valori diferite pentru $\overline{xy}$, cu excepţia cazului examinat la punctul 3 al acestui capitol, cînd variabilele aleatoare considerate sînt independente.

În acest caz dacă una din variabile ia o valoare determinată, acest rezultat nu influenţează legea de probabilitate a celei de-a doua variabile aleatoare.

Scanned with OKEN Scanner

Capitolul V

Dispersie. Abateri medii

1. Considerații generale

În practică se întîmplă adeseori ca valoarea medie a unei variabile aleatoare să nu ne ofere decît o imagine aproximativă, orientativă asupra caracterului variabilei studiate, rămînînd necunoscute trăsăturile ei esențiale. De aceea este necesar să se cunoască mai amănunțit legea de probabilitate respectivă, încercîndu-se calculul a încă unul sau doi indicatori de același gen ca și valoarea medie, astfel încît acest mic grup de indicatori să ne dea în ansamblu o caracterizare suficientă a acelor trăsături esențiale ale variabilei considerate care interesează în practică.

De exemplu, din punct de vedere practic, interesează, îndeosebi, problema abaterii valorilor efective ale variabilei aleatoare date, de la valoarea ei medie; interesează, deci, cît de mult sînt *împrăștiate* sau „dispersate" toate aceste valori efective.

Sînt posibile legi de probabilitate a căror valori, în majoritatea lor, vor fi foarte apropiate de valoarea lor medie, deci și foarte apropiate între ele, sau, dimpotrivă, legi de probabilitate cu valori în majoritatea lor foarte împrăștiate, simțitor diferite de valoarea lor medie; în acest caz valorile vor diferi foarte mult și între ele.

Mai mult, sînt posibile legi de probabilitate care au aceeași valoare medie a variabilei aleatoare și totuși au amplitudini ale variației valorilor ca și dispersiile lor în jurul mediilor respective, sensibil diferite.

Să presupunem că se face proba productivității a două soiuri de grîu *). Se știe că recolta la un metru patrat este supusă de obicei unor variații simțitoare, în funcție de anumite circumstanțe întîmplătoare cum sînt: cantitatea precipitațiilor, repartiția îngrășămintelor etc. În proba

*) Vezi B. V. Gnedenko și A. I. Hincin, *Introducere elementară în calculul probabilităților*. Ed. Tehnică, București, 1953.

considerată să presupunem că pentru cele două soiuri de grîu s-au asigurat condiţii identice de dezvoltare şi că s-a obţinut aceeaşi recoltă medie şi anume 240 g la un metru pătrat. Este evident că în acest caz indicatorul valorii medii nu este suficient pentru a ne caracteriza variabila aleatoare considerată (recolta la un metru pătrat), deoarece, din punct de vedere economic, un mare interes îl prezintă acel soi de grîu a cărei productivitate este mai puţin supusă influenţelor întîmplătoare ale factorilor meteorologici şi ale altor factori; interesează aşadar soiul de grîu care are *dispersia* productivităţii mai mică.

Problema care se pune este aceea de a se găsi un indicator care să constituie, în mod raţional, o măsură a dispersiei unei variabile aleatoare. Un asemenea indicator trebuie să ne arate la ce abateri de la valoarea medie ne putem aştepta în cazul variabilei considerate.

În capitolul precedent s-a arătat (IV,1.3) că abaterea

$$\xi = X - M(X)$$

este şi ea o variabilă aleatoare care-şi are legea ei de probabilitate şi că avem teorema $M(\xi) = 0$. Se vede, deci, că valoarea medie a abaterii nu ne poate servi drept măsură a împrăştierii — a dispersiei valorilor variabile aleatoare X în jurul valorii ei medii $M(X)$.

Este firesc însă să considerăm *abaterea medie* în valoare absolută, sau *abaterea medie pătratică*, drept indicatori proprii pentru a caracteriza mai complet variabila aleatoare luată în studiu.

În teoria calculului probabilităţilor se cunosc şi alţi indicatori ai împrăştierii, însă în cele ce urmează nu ne vom opri decît asupra celor doi amintiţi mai sus şi îndeosebi asupra indicatorului abaterii medii pătratice, deoarece posedă mai multe însuşiri speciale care o fac să fie preferabilă tuturor celorlalte măsuri ale abaterii.

2. Abatere medie

Abaterile valorilor unei variabile aleatoare pot fi măsurate, după cum s-a mai spus, cu media abaterilor absolute, $|X - M(X)|$.

Dacă revenim la exemplul 22 şi vrem să caracterizăm mai profund calitatea cunoştinţelor de matematici ale elevilor din două şcoli medii, în acest scop putem calcula media abaterii în valoare absolută, ale celor două variabile şi să le comparăm între ele.

Notînd prin

$$|X_i - M(X)| = |\xi_i|$$

abaterea în valoare absolută a variabilelor X_i, avem următoarea lege de probabilitate a abaterii:

$$|\xi_i| = |X_i - M(X)| = \begin{bmatrix} |1-21{,}1|, & |10-21{,}1|, & |20-21{,}1|, & |30-21{,}1| \\ 0{,}1 & 0{,}1 & 0{,}4 & 0{,}4 \end{bmatrix}.$$
$$p_i$$

Calculînd media acestei variabile pe care o notăm cu $M(\xi)$, avem

$$M(\xi) = (20{,}1 \cdot 0{,}1) + (11{,}1 \cdot 0{,}1) + (1{,}1 \cdot 0{,}4) + (8{,}9 \cdot 0{,}4) = 7{,}12.$$

În mod analog, notînd cu

$$| Y_i - M(Y) | = (\xi_i)$$

abaterea în valoare absolută a valorilor Y_i, vom avea legea de probabilitate a acestei variabile:

$$| \eta_i | = Y_i - M(Y) | = \begin{bmatrix} | 1-21{,}0 |, & | 10-21{,}0 |, & 20-21{,}0 |, & | 30-21{,}0 | \\ 0 & 0{,}2 & 0{,}5 & 0{,}3 \end{bmatrix}.$$
$$p_i \quad =$$

Media acestei variabile pe care o notăm cu $M(\eta)$ este $M(\eta)$: $(20{,}0 \cdot 0) + (11{,}0 \cdot 0{,}2) + (1{,}0 \cdot 0{,}5) + (9 \cdot 0{,}3) = 5{,}4$.

Observăm că pentru variabila Y, abaterea medie în valoare absolută este mult mai mică decît pentru variabila X. Din punct de vedere practic aceasta înseamnă, evident, că, cu toate că în medie, atît elevul din școala A cît și elevul din școala B realizează cam același număr de puncte (am văzut că valorile medii ale numărului de puncte obținute sînt respectiv 21,1 și 21,0), totuși, numărul de puncte obținut de răspunsul unui elev din B este mai omogen, el variază de la 10 la 30 puncte pe cînd numărul de puncte obținute de elevii din A este mai împrăștiat variind între 1 și 30 puncte. Această trăsătură caracteristică a celor două legi de probabilitate se reflectă și în nivelul indicatorilor $M(\xi)$ și $M(\eta)$.

Menționăm că în cercetările teoretice calculul cu mărimi absolute dă naștere la complicații analitice, fapt care a determinat folosirea aproape exclusivă a abaterii medii pătratice drept măsură a concentrării abaterilor.

3. Dispersie. Abaterea medie pătratică

La fel ca și abaterile $x_i - \bar{x}$ ale variabilei aleatoare X de la valoarea ei medie (X), pătratul $(x-\bar{x})^2$ este o variabilă aleatoare a cărei lege de probabilitate este

$$\xi^2 = \begin{bmatrix} (x_1-\bar{x})^2, & (x_2-\bar{x})^2, & \ldots, & (x_n-\bar{x})^2 \\ p_1, & p_2, & \ldots, & p_n \end{bmatrix}. \qquad \text{(V,3.1)}$$

Valoarea medie a variabilei aleatoare ξ^2 este deci

$$M(\xi^2) = \sum_{i=1}^{n} (x_i - \bar{x})^2 p_i = \sigma^2 = D^2(X). \qquad \text{(V,3.2)}$$

Această mărime, notată de obicei prin $\sigma^2 = D^2(X)$, se numește *dispersia* variabilei aleatoare X. Dispersia ne dă o idee aproximativă asupra distribuției pătratului abaterii $X - M(X)$. Extrăgînd rădăcina pătrată din valoarea indicatorului (V,3.2), obținem o altă valoare care poate caracteriza nivelul aproximativ al abaterii însăși. Această valoare

se numeşte *abaterea medie pătratică* şi se notează, în general, cu μ sau σ:

$$\mu = \sigma = \sqrt{M(\xi^2)}. \qquad (V,3.3)$$

Ţinînd seama de (V.3.2), obţinem

$$\mu = \sqrt{\Sigma\, p_i\,(x_i - \overline{x})^2}\,; \qquad (V,3.4)$$

efectuînd calculele, rezultă

$$\mu = \sqrt{\Sigma p_i x_i^2 - 2\overline{x}\ \Sigma p_i x_i + \overline{x}^2 \Sigma p_i}\,. \qquad (V,3.5)$$

Întrucît

$$\Sigma p_i = 1\,;\ \Sigma p_i x_i = M(X)\,;\ \Sigma p_i x_i^2 = M(X^2),$$

abaterea medie pătratică μ se mai poate scrie, făcînd înlocuirile în (V,3.5), sub forma

$$\mu = \sqrt{M(X^2) - [M(X)]^2}. \qquad (V,3.6)$$

În afară de aceste avantaje, folosirea acestui indicator drept măsură a abaterii diferitelor valori ale unei variabile aleatoare de la valoarea lor medie, simplifică simţitor calculele practice.

Calculînd abaterile medii pătratice pentru legile de probabilitate (IV,8.1) şi (IV,8.2) din exemplul 22, obţinem, pentru variabila X,

$$\mu_x = \sqrt{[(-20{,}1)^2 \cdot 0{,}1] + [(-11{,}1)^2 \cdot 0{,}1] + [(-1{,}1)^2 \cdot 0{,}4] + [(8{,}9)^2 \cdot 0{,}4]} =$$

$$\sqrt{84{,}890} = 9{,}218,$$

iar pentru variabila Y

$$\mu_y = \sqrt{[(-11)^2 \cdot 0{,}2] + [(-1)^2 \cdot 0{,}5] + [(9)^2 \cdot 0{,}3]} = \sqrt{49} = 7.$$

Observăm că şi abaterea medie pătratică a variabilei Y este mai mică decît abaterea medie pătratică a variabilei X, ceea ce înseamnă că numărul de puncte pe care-l obţin elevii din şcoala B este concentrat mai omogen în jurul valorii medii, deşi nivelul valorii medii la elevii din şcoala A este mai mare.

4. Abaterea medie pătratică a numărului de bile albe obţinute în n extracţii în cazul urnei Bernoulli

După cum ştim, dacă notăm cu X variabila noastră aleatoare, dată de numărul de bile albe extrase în n probe, legea ei de probabilitate este următoarea:

$$X = \begin{bmatrix} 0, & 1, & 2, & \dots & n \\ q^n, & C_n^1 q^{n-1} p, & C_n^2 q^{n-2} p^2, & \dots & p^n \end{bmatrix}$$

în care $q = 1 - p$.

Din relația precedentă deducem legea de probabilitate a variabilei aleatoare ξ, dată de abaterea de la valoarea medie:

$$\xi\begin{bmatrix} 0-np, & 1-np, & 2-np, & \dots & n-np \\ q^n, & C_n^1 q^{n-1}p, & C_n^2 q^{n-2}p^2, & \dots & p^n \end{bmatrix}.$$

Știm că

$$\xi = \xi_1 + \xi_2 + \dots + \xi_n = \sum_{i=1}^{n} \xi_i,$$

unde $\xi_i = X_i - p$, iar X_i ia valorile 1 sau 0 după cum la extracția de rangul i, evenimentul A de a scoate o bilă albă are loc sau nu are loc. Dacă ξ_i reprezintă abaterea privind extracția de rangul i și ia valorile $1-p$ și $0-p = -p$, cu probabilitățile respective p și q, vom avea legea de probabilitate a variabilei ξ_i:

$$\xi_i\begin{bmatrix} 1-p & -p \\ p & q \end{bmatrix}.$$

Valoarea medie pătratică (dispersia) pentru această lege de probabilitate este

$$M(\xi_i^2) = p(1-p)^2 + q(-p)^2 = pq(p+q) = pq \cdot 1 = pq.$$

Deoarece în baza teoremei cunoscute (IV, 1.3), $M(\xi) = 0$ și variabilele $\xi_1, \xi_2, \dots, \xi_n$ sînt independente, putem considera, ținînd seama de (IV, 6.1),

$$M(\xi^2) = M[(\xi_i + \xi_2 + \dots + \xi_n)^2] = M(\xi_1^2) + M(\xi_2^2) \dots + M(\xi_n^2)$$

și întrucît

$$M(\xi_i^2) = pq,$$

obținem

$$M(\xi^2) = npq \qquad (V,4.1)$$

care reprezintă tocmai dispersia abaterii valorilor variabilei X de la valoarea lor medie. Din (V,4.1) se obțin, după cum știm, *abaterea medie pătratică*

$$\mu = \sigma = \sqrt{npq}. \qquad (V,4.2)$$

Mai departe ne vom folosi adeseori de această formulă la studiul fenomenelor care se apropie de schema urnei Bernoulli.

5. Abaterea medie pătratică a unei sume de variabile aleatoare

Dacă presupunem că avem n variabile aleatoare independente $x_1, x_2, \dots, x_n$ cu abaterile medii pătratice $\mu_1, \mu_2, \dots, \mu_n$ și că

$$\Sigma x_i = X$$

pentru care μ este abaterea medie pătratică, avem următoarea regulă de însemnătate fundamentală în diferite aplicații :

Dispersia sumei unor variabile aleatoare independente între ele este egală cu suma dispersiilor lor.

Avem deci

$$\mu^2 = \mu_1^2 + \mu_2^2 + \ldots + \mu_n^2 = \sum_{i=1}^{n} \mu_i^2 \qquad (V,5.1)$$

Într-adevăr, din relația

$$X = x_1 + x_2 + \ldots + x_n = \sum_{i=1}^{n} x_i$$

deducem că

$$M(X) = M(x_1) + M(x_2) + \ldots + M(x_n)$$

de unde, determinînd abaterile, obținem

$$X - M(X) = x_1 + x_2 + \ldots + x_n - M(x_1) - M(x_2) - \ldots - M(x_n) =$$
$$= x_1 - M(x_1) + x_2 - M(x_2) + \ldots + x_n - M(x_n);$$

deci

$$M[X - M(X)]^2 = M\ \Sigma[x_i - M(x_i)]^2. \qquad (V,5.2)$$

Dezvoltînd partea a doua a egalității (V,5.2), obținem

$$M[X - M(X)]^2 = \sum_{i=1}^{n} M[x_i - M(x_i)]^2 + 2\sum_{\substack{i=1 \\ i \neq k}}^{n} \sum_{k=1}^{n} M[x_i - M(x_i)]\ M[x_k - M(x_k)]. \qquad (V,5.3)$$

Observăm că în relația (V,5.3) putem introduce notația la care am convenit mai sus, adică

$$M[X - M(X)]^2 = \sigma^2,$$

$$\Sigma M[x_i - M(x_i)]^2 = \sum_{i=1}^{n} \sigma_i^2,$$

$$\sum_{i=1}^{n} \sum_{k=1}^{n} M[x_i - M(x_i)] \cdot M[x_k - M(x_k)] = \sum_{i=1}^{n} \sum_{k=1}^{n} M(x_i - \bar{x}_i) \cdot M(x_k - \bar{x}_k).$$

Prin urmare, relația (V,5.3) se transformă în relația

$$\sigma^2 = \sum_{i=1}^{n} \sigma_i^2 + \sum_{k=1}^{n} \sum_{\substack{k=1 \\ i \neq k}}^{n} M(x_i - \bar{x}_i)\ M(x_k - \bar{x}_k). \qquad (V,5.4)$$

Însă, se poate observa că dubla sumă din partea a doua a egalității este egală cu zero, pentru că cei doi factori sînt nuli datorită faptului că variabilele sînt independente între ele.

Într-adevăr,

$$M(x_i - \bar{x}_i) = M(x_i) - M(\bar{x}_i) = 0.$$

Prin urmare din relația (V, 5.4) obținem

$$\sigma^2 = \sum_{i=1}^{n} \sigma_i^2 = \sigma_1^2 + \sigma_2^2 + \ldots + \sigma_n^2. \qquad (V,5.5)$$

Scanned with OKEN Scanner

Este evident că această regulă de adunare a dispersiilor permite un calcul simplu și pentru abaterile medii pătratice. Vom avea, prin urmare,

$$\mu = \sqrt{\sum_{i=1}^{n} \mu_i^2}, \qquad (V,5.6)$$

Această regulă, care ne oferă posibilitatea de a exprima simplu abaterea medie pătratică a sumei prin abaterile medii pătratice ale termenilor sumei, în cazul evenimentelor independente, constituie unul din principalele avantaje ale indicatorului abaterii medii pătratice față de ceilalți indicatori de măsură a concentrării abaterilor.

EXEMPLUL 24. Să presupunem că rezultatele a n măsurători ale unei mărimi oarecare ne-au dat valorile $X_1, X_2, \ldots, X_n$, că $\overline{X}$ este valoarea medie pentru fiecare din aceste variabile aleatoare, iar μ este abaterea medie pătratică ce rămîne constantă la toate măsurătorile, deoarece ele se fac în condițiile aceleiași legi de probabilitate. Variabilele $X_1, X_2, \ldots, X_n$ sînt independente între ele.

Ni se cere să calculăm valoarea medie și abaterea medie pătratică a tuturor măsurătorilor luate în ansamblu și să interpretăm rezultatul obținut.

Este evident că în baza teoremei (IV,2.3), cu privire la valoarea medie a sumei de variabile independente, ajungem la concluzia că

$$M_1 = \frac{1}{n} M(X_1 + X_2 + \ldots + X_n) = \frac{1}{n}(\overline{X}_1 + \overline{X}_2 + \overline{X}_3 + \ldots + \overline{X}_n) =$$

$$= \frac{1}{n}(n\overline{X}) = \overline{X}$$

ceea ce înseamnă că valoarea medie a tuturor măsurătorilor este aceeași ca la fiecare măsurătoare în parte.

Considerînd, mai departe, regula de adunare a dispersiilor (V,5.5), abaterea medie pătratică a sumei $X_1 + X_2 + \ldots + X_n$ rezultă egală cu

$$\mu(X_1 + X_2 + \ldots + X_n) = \sqrt{n\mu^2} = \mu\sqrt{n}. \qquad (V,5.7)$$

Abaterea medie pătratică a variabilei

$$Y = \frac{X_1 + \ldots + X_n}{n}$$

rezultă imediat din egalitățile

$$\mu_y = \frac{\mu(X_1 + X_2 + \ldots + X_n)}{n} = \frac{\mu\sqrt{n}}{n} = \frac{\mu}{\sqrt{n}}. \qquad (V,5.8)$$

Ce ne spune acest rezultat?

El ne arată că abaterea medie pătratică a mediei aritmetice a unui număr de variabile independente este mai mică decît abaterea medie pătratică a fiecărei variabile în parte (este de $\sqrt{n}$ ori mai mică).

6. Momente centrate

Prin *momente centrate* înțelegem valorile medii ale puterilor de diferite ordine ale abaterilor valorilor unei variabile aleatoare de la valoarea lor medie. Astfel, dispersia

$$\sigma^2 = D^2(X) = \sum_{i=1}^{n} (x_i - \overline{x})^2 p_i,$$

este momentul centrat de ordinul doi.

În general, momentul centrat de ordinul k este dat de relația

$$\mu_k = \sum_{i=1}^{n} p_i [(x_i - M(x)]^k . \qquad (V,6.1)$$

Momentele centrate reprezintă anumite proprietăți intrinseci ale variabilelor aleatoare. În particular, am văzut proprietățile analitice ale dispersiei.

Cunoașterea, îndeosebi, a primelor patru momente ale unei legi de probabilitate prezintă interes practic și incontestabil cînd sîntem puși în situația de a caracteriza diferite distribuții statistice din natură sau din societate.

Asupra interpretării diferitelor momente centrate ale legilor de probabilitate pentru o singură variabilă sau pentru un sistem de variabile independente vom reveni mai departe.

7. Inegalitatea lui Cebîșev *)

Am văzut pînă acum că studiul unuia din momentele centrate — studiul dispersiei, cu ajutorul abaterii medii pătratice — ne-a dat o primă indicație asupra mărimii pe care o pot lua abaterile valorilor unei variabile aleatoare. Totuși, abaterea medie pătratică nu ne permite să facem o apreciere exactă asupra numărului aproximativ al abaterilor mari, în sens pozitiv sau negativ. Cu alte cuvinte, cunoașterea abaterii medii pătratice ne indică doar gradul mai mare sau mai mic de concentrare a valorilor date ale variabilei în jurul valorii lor medii; ea nu ne dă o indicație precisă asupra mărimii probabilității de a obține abateri de o mărime

*) P. L. Cebîșev (1821—1894), ilustru matematician rus, fondatorul școalei matematice din Petersburg, din care au făcut parte valoroși matematicieni ca A.A. Markov, A. M. Liapunov, E. I. Zolotarev și alții.

Opera lui Cebîșev este vastă. El a adus contribuții însemnate în teoria numerelor, calculul probabilităților, teoria aproximării funcțiilor prin polinoame, teoria mecanismelor etc. Cebîșev a reușit să obțină rezultate științifice de mare importanță prin mijloace simple, elementare (a se vedea ca exemplu inegalitățile lui Cebîșev). În calculul probabilităților a demonstrat teorema centrală prin metoda momentelor, construind o teorie generală a polinoamelor ortogonale dintre care unele îi poartă numele.

A aplicat matematica în diverse domenii practice, fiind astfel unul din precursorii matematicelor aplicate.

dată. Marele matematician rus Cebîșev a stabilit însă o limită a acestei probabilități printr-o inegalitate care-i poartă numele. Mai tîrziu, francezul Bienaymé a găsit aceeași inegalitate, fără să cunoască lucrările lui Cebîșev.

Inegalitatea lui Cebîșev se poate formula în două moduri echivalente :

$$P[|X - M(X)| \geqslant L] \leqslant \frac{\sigma^2}{L^2} \qquad (V,7.1)$$

și

$$P[|X - M(X)| < \mathbf{L}] \geqslant 1 - \frac{\sigma^2}{L^2} \qquad (V,7.2)$$

în care L este un număr pozitiv oarecare, iar ceilalți indicatori ne sînt cunoscuți.

Inegalitatea (V,7.1) ne spune că *probabilitatea ca abaterile în valori absolute să fie mai mari decît un număr pozitiv oarecare L este mai mică decît dispersia împărțită la L^2.*

Același lucru, însă sub o altă formă, ne spune inegalitatea (V,7.2); în această inegalitate se afirmă că *probabilitatea ca abaterile în valori absolute să fie mai mici decît un număr pozitiv oarecare L este mai mare decît $1 - \frac{\sigma^2}{L^2}$.*

Este vorba, deci, de o limită inferioară a probabilităților ca abaterile în valori absolute ale valorilor unei variabile aleatoare X să fie mai mici decît un număr pozitiv oarecare.

Dacă luăm drept unitate de măsură pentru abateri, abaterea medie pătratică, $\sigma = \mu$, și evaluăm pe L cu ajutorul acestei unități de măsură, obținem $L = k\sigma$, în care k este un coeficient oarecare; prin urmare inegalitatea (V,7.2) poate fi scrisă și în felul următor :

$$P[|X - M(X)| \leqslant k\sigma] \geqslant 1 - \frac{\sigma^2}{k^2\sigma^2} = 1 - \frac{1}{k^2}. \qquad (V,7.2')$$

Să dăm lui k diferite valori.

Astfel dacă $k = 2$, avem

$$P[|X - M(X)| \leqslant 2\sigma] \geqslant 1 - \frac{\sigma^2}{2^2\sigma^2} = \frac{3}{4}.$$

Aceasta înseamnă că — cu o probabilitate cuprinsă între $\frac{3}{4}$ și 1 — majoritatea abaterilor valorilor oricărei variabile aleatoare nu va depăși 2σ; *adică în medie mai mult de 75% din numărul abaterilor (în valori absolute) vor fi mai mici decît 2σ.*

Un calcul simplu ne arată că dacă avem $k = 7$, atunci mai mult de 98% din numărul abaterilor (în valori absolute), vor fi mai mici, în general, decît 7σ.

Să luăm exemplul clasic al aruncării unei monede.

Scanned with OKEN Scanner

EXEMPLUL 25. Aruncăm o monedă de 100 ori. Sîntem în cazul schemei probabilistice a unei urne Bernoulli cu două stări, în care $p = \frac{1}{2}$; $q = \frac{1}{2}$.

Ni se cere să arătăm care este limita inferioară a probabilității ca să obținem abateri absolute, mai mici decît 3σ, ale valorilor pe care le poate lua apariția uneia din cele două fețe.

Pentru ca să putem aplica inegalitatea lui Cebîșev în forma (V,7.2), trebuie să calculăm media și abaterea medie pătratică a variabilei noastre aleatoare care reprezintă numărul de apariții, teoretic posibile, ale feței care ne interesează. Dacă notăm cu X variabila noastră, ea admite următoarea lege de probabilitate :

$$X\begin{bmatrix} 0, & 1, & 2, & \ldots, & n \\ q^n, & C_n^1 q^{n-1} p, & C_n^2 q^{n-2} p^2, & \ldots, & p^n \end{bmatrix}.$$

Conform relației (IV,4.3) știm că valoarea medie a legii de probabilitate pentru urna Bernoulli este

$$M(X) = np,$$

iar conform relației (V. 4.2)

$$\sigma = \sqrt{npq}.$$

Aflăm deci că, în exemplul nostru :

$$M(X) = np = 100 \cdot \frac{1}{2} = 50,$$

$$\sigma^2 = npq = 100 \cdot \frac{1}{2} \cdot \frac{1}{2} = 25,$$

$$\sigma = \sqrt{npq} = \sqrt{25} = 5.$$

Avînd aceste date, le introducem în inegalitatea (V,7.2) și obținem

$$P[|X - M(X)| \leqslant 3.5] \geqslant 1 - \frac{1}{3^2}.$$

Sau, notînd

$$\alpha = X - M(X)$$

înseamnă că numărul cazurilor în care α se va abate de la medie în plus sau în minus, cu cel mult $15 = 3\sigma$ va fi obținut cu probabilitatea :

$$P[35 \leqslant \alpha \leqslant 65] \geqslant 1 - \frac{1}{9} \geqslant \frac{8}{9} = 0{,}88 = 88\%.$$

Prin urmare, cel puțin 88% din numărul valorilor obținute, vor fi cuprinse, în medie, între valorile 35 și 65.

Exprimînd această probabilitate prin frecvențe relative, vom avea

$$P[0{,}35 \leqslant \frac{\alpha}{100} \leqslant 0{,}65] \geqslant 1 - \frac{1}{9} = \frac{8}{9} = 88\%.$$

Dacă în același exemplu considerăm $n = 10\,000$ și $k = 6$, vom avea

$$M(X) = 5000.$$
$$\sigma^2 = 2\,500,$$
$$\sigma = \sqrt{2\,500} = 50,\ k\sigma = 6 \cdot 50 = 300,$$

iar limita inferioară a probabilității în baza relației (V,7.2):

$$P\left[\frac{4\,700}{10\,000} \leqslant \frac{\alpha}{10\,000} \leqslant \frac{5\,300}{10\,000}\right] \geqslant 1 - \frac{1}{6^2} = \frac{35}{36} = 0{,}9722.$$

Observăm că în cel puțin 9 722 de cazuri din 10 000, abaterile valorilor față de media teoretică de 50% vor fi foarte mici; ele vor fi de maximum 3% în plus sau în minus, față de limitele de mai sus.

Ultima ipoteză ne duce la constatarea că *cu cît numărul probelor crește, cu atît frecvența se apropie de probabilitatea teoretică.*

Inegalitatea lui Cebîșev se poate demonstra foarte simplu, pornind de la dispersia următoarei legi de probabilitate

$$X_i \begin{bmatrix} x_1, & x_2, & \ldots, & x_n \\ p_1, & p_2, & \ldots, & p_n \end{bmatrix}.$$

Vom avea deci

$$\sigma^2 = [X_1 - M(X)]^2 p_1 + [(X_2 - M(X)]^2 p_2 + \ldots + [(X_n - M(X)]^2 p_n. \qquad \text{(V,7.3)}$$

Notăm cu α_i abaterile $X_i - M(X)$ — pentru ușurința calculului.

Vom avea atunci

$$\sigma^2 = p_1 \alpha_1^2 + p_2 \alpha_2^2 + \ldots + p_n \alpha_n^2. \qquad \text{(V,7.4)}$$

Valorile α_i le aranjăm în ordinea mărimii lor crescătoare, astfel ca

$$\alpha_1^2 \leqslant \alpha_2^2 \leqslant \ldots \leqslant \alpha_n^2. \qquad \text{(V,7.5)}$$

Să considerăm acum un număr pozitiv oarecare L care împarte inegalitățile (V,7.5) în două grupe de valori, unele α_i^2 sînt mai mici ca L, iar altele $\alpha_{i+1}^2, \alpha_{i+2}^2, \ldots, \alpha_n^2$ sînt mai mari sau egale cu L, abaterile $\alpha_1, \alpha_2, \ldots, \alpha_i$ și $\alpha_{i+1}, \ldots, \alpha_n$ luîndu-se în valori absolute. Inegalitățile (V,7.5) se vor putea scrie în acest caz:

$$|\alpha_1^2| \leqslant |\alpha_2^2| \leqslant \ldots \leqslant |\alpha_i^2| \leqslant L \leqslant |\alpha_{i+1}^2| \leqslant |\alpha_{i+2}^2| \leqslant \ldots \leqslant |\alpha_n^2|.$$

Dacă din suma (V, 7.4) eliminăm toți termenii mai mici ca L egalîndu-i cu zero și lăsăm numai termenii mai mari sau egali cu L egalîndu-i cu L^2, atunci (V, 7.4) se va putea scrie sub forma inegalității

$$\sigma^2 \geqslant L^2(p_{i+1} + p_{i+2} + \ldots + p_n), \qquad \text{(V,7.6)}$$

de unde

$$p_{i+1} + p_{i+2} + \ldots + p_n \leqslant \frac{\sigma^2}{L^2}. \qquad \text{(V,7.7)}$$

Observăm că prima parte a inegalității (V, 7.7) reprezintă — în baza regulei de adunare a probabilităților — probabilitatea ca să avem o abatere mai mare sau egală cu numărul pozitiv L, ceea ce înseamnă că ea este probabilitatea evenimentului complementar ca să avem o abatere mai mică decît L, adică

$$p_{i+1} + p_{i+2} + \ldots + p_n = 1 - (p_1 + p_2 + \ldots + p_i).$$

Înlocuind în inegalitatea (V,7.6), obținem

$$\sigma^2 \geqslant \left(1 - \sum_1^i p_i\right) L^2$$

sau

$$\frac{\sigma^2}{L^2} \geqslant 1 - \sum_1^i p_i$$

și prin urmare

$$\sum_1^i p_i \geqslant 1 - \frac{\sigma^2}{L^2}.$$

După cum am mai arătat, $\sum_1^i p_i$ reprezintă probabilitatea ca x să ia una dintre valorile $x_1, x_2, \ldots, x_i$, adică probabilitatea ca să obținem pentru x o abatere mai mică decît L:

$$P[|x_i - M(X)| < L] \geqslant 1 - \frac{\sigma^2}{L^2},$$

care este tocmai inegalitatea lui Cebîșev (vezi relația (V, 7.2).

Atît pentru calculul probabilităților cît și pentru statistica matematică, inegalitatea lui Cebîșev este fundamentală. Este evident că această inegalitate nu este singura pe care o putem construi. De exemplu, noi am fi putut îmbunătăți limita inferioară a probabilității, luînd momente centrate de alt ordin. Necesitatea de a lua o altă inegalitate, impusă de scopuri practice, care nu se bazează pe dispersie, s-a ivit în domeniul asigurărilor, la calculul riscului de asigurare. Se știe că în acest domeniu, pentru a delimita limita inferioară a probabilității se folosește momentul centrat de ordinul patru al variabilei aleatoare considerate.

Se ajunge astfel la inegalitatea

$$P[|X - M(X)| < L] \geqslant 1 - \frac{\mu_4}{L^4},$$

cunoscută sub numele de inegalitatea lui Medolaghi.

Pentru demonstrarea acestei formule, vom folosi următoarea teoremă, mai generală, datorită tot lui Cebîșev.

TEOREMA lui CEBÎȘEV. *Fie $F(x)$ o funcție nenegativă de variabila aleatoare X. Pentru oricare număr $K > 0$, avem inegalitatea*

$$P[F(X) \geqslant K] \leqslant \frac{M[F(X)]}{K}.$$

Fie $x_1, x_2, \ldots, x_n$ valorile pe care le poate lua X, cu probabilitățile respective $p_1, p_2, \ldots, p_n$ $\left(\sum_{i=1}^n p_i = 1\right)$.

Fie I mulțimea indicilor pentru care avem

$$F(x_i) \geqslant k, \quad i \in I.$$

Evident,

$$I \subset [1, 2, \ldots, n].$$

Avem

$$M[F(X)] = \sum_{i=1}^n p_i F(x_i) \geqslant \sum_{i \in I} p_i F(x_i) \geqslant K \sum_{i \in I} p_i.$$

Însă

$$\sum_{i \in I} p_i = P[F(X) \geqslant K].$$

Din aceste două relații rezultă imediat formula lui Cebîșev, enunțată în teoremă. Dacă luăm

$$F(X) = [X - M(X)]^4, \; K = L^4,$$

formula lui Cebîșev ne dă

$$P[[X - M(X)]^4 \geqslant L^4] \leqslant \frac{M[X - M(X)]^4}{L^4}$$

sau

$$P[|X - M(X)| \geqslant L] \leqslant \frac{\mu_4}{L^4}.$$

Cum

$$P[|X - M(X)| < L] = 1 - P[|X - M(X)| \geqslant L,$$

din inegalitatea de mai sus rezultă imediat inegalitatea lui Medolaghi.

Scanned with OKEN Scanner

Scanned with OKEN Scanner

Capitolul VI

Funcţie de repartiţie. Funcţie caracteristică

1. Funcţie de repartiţie. Densitate de probabilitate

Am văzut că orice lege de probabilitate se caracterizează prin faptul că oricărei valori pe care poate s-o ia variabila aleatoare, i se asociază o probabilitate corespunzătoare.

Astfel, în cazul urnei Bernoulli compusă din n_1 bile albe şi n_2 bile negre, dacă atribuim valoarea 0 oricărei bile albe extrase şi valoarea 1, oricărei extracţii în care ne apare o bilă neagră, este evident că valorii 0 îi vom asocia probabilitatea

$$p = \frac{n_1}{n_1 + n_2}$$

de a extrage o bilă albă, iar valorii 1 îi vom asocia probabilitatea

$$p = \frac{n_2}{n_1 + n_2}$$

de a extrage o bilă neagră.

Este util să precizăm acum că orice lege de probabilitate poate fi reprezentată grafic cu ajutorul diagramelor care ne vor permite să desprindem cu uşurinţă trăsăturile principale ale legilor reprezentate. Schema unei asemenea reprezentări este dată de o dreaptă orizontală pe care sînt dispuse toate valorile posibile ale variabilei aleatoare X, începînd de la o origine O oarecare, valorile pozitive fiind desigur spre dreapta, iar cele negative spre stînga.

În punctele dreptei care reprezintă diferite valori ale variabilei aleatoare X, se ridică verticale care — la o anumită scară — reprezintă probabilităţile valorilor respective (vezi fig. 1).

O scurtă privire asupra figurii ne arată că variabila aleatoare cu cea mai mare probabilitate se găsește în punctul x_6; aceasta e cea mai probabilă valoare. Se observă de asemenea că pe măsură ce valorile posibile se îndepărtează de valoarea cea mai probabilă (x_6), probabilitățile lor scad necontenit.

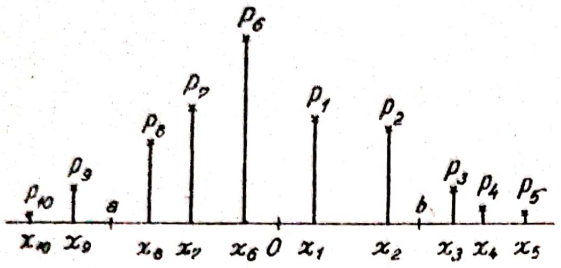


Fig. 1

Dacă considerăm intervalul (a, b) și ne punem întrebarea care este probabilitatea ca variabila noastră aleatoare să ia o valoare cuprinsă în acest interval, în baza regulii de adunare a probabilităților, probabilitatea căutată va fi egală cu suma probabilităților tuturor valorilor posibile situate în interval. Ea va fi reprezentată de suma lungimilor liniilor verticale aflate în interval.

În figura 1 vom avea

$$P(a < x < b) = p_1 + p_2 + p_6 + p_7 + p_8 .$$

În practica statistică, valorile pe care le poate lua o variabilă aleatoare pot fi în număr finit sau să formeze un interval finit sau infinit.

În primul caz avem o lege de probabilitate *discontinuă*. În cazul al doilea, întîlnim în practica statistică în general, legi *continue*, caracterizate prin faptul că reprezentarea grafică a legii de probabilitate corespunzătoare, făcută în mod analog ca în graficul din figura 1, este dată de o funcție continuă.

Să presupunem, de exemplu, că avem o variabilă aleatoare cu un număr finit de valori $x_i, (i = 1,2,3,\ldots, n)$. Reprezentînd grafic valorile probabilităților, ele vor avea forma unor arii dreptunghiulare cu baza egală cu unitatea (distanța între două valori vecine) iar înălțimea egală cu linia verticală reprezentînd probabilitatea respectivă; aria dreptunghiului

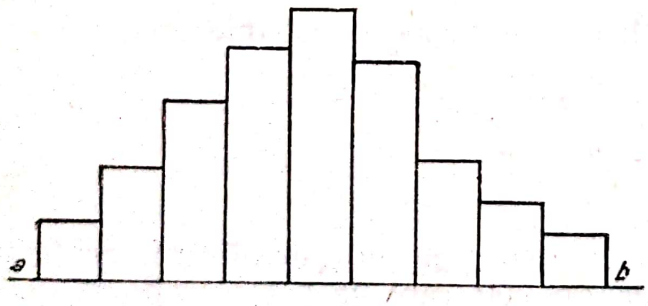


Fig. 2

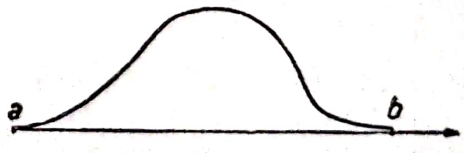


Fig. 3

este egală chiar cu lungimea liniei verticale. Astfel probabilitatea inegalităților $a < x < b$ este reprezentată de suma ariilor dreptunghiurilor așezate deasupra intervalului (a, b) (vezi fig. 2).

În cazul unei legi de probabilitate continue, reprezentarea grafică a funcției de probabilitate într-un interval dat se face printr-o curbă continuă (vezi fig. 3).

Este evident că determinarea probabilității unei valori egale cu un număr dat își pierde în acest caz orice interes practic. Un interes esențial îl prezintă *probabilitatea ca o valoare oarecare să fie cuprinsă într-un interval dat*. De pildă, un interes practic îl prezintă probabilitatea ca recolta medie de grîu la hectar să fie cuprinsă între 2 500 și 2 700 kg, în cazul cînd se evaluează recolta probabilă într-o gospodărie colectivă oarecare. Această probabilitate este dată de suprafața ariei delimitată în jos de intervalul (a, b) considerat, iar în sus de curba reprezentativă a probabilității.

Inegalitatea $0 \leqslant p \leqslant 1$, în reprezentare grafică, trebuie interpretată în sensul că suprafața ariei este cuprinsă între 0 și 1.

Legile de probabilitate studiate pînă acum se referă la valori luate, pe axa reală. De aceea, ele poartă numele de legi unidimensionale.

Am examinat pînă acum cazul unei legi de probabilitate, în care evenimentele sînt în număr finit.

Am mai dat și exemplul unei legi de probabilitate, care poate lua toate valorile cuprinse într-un interval dat și a cărei reprezentare este dată de o curbă continuă.

Aceste legi, frecvent întîlnite în statistică, sînt esențial diferite. Totuși ele pot să fie înglobate în teoria lui Kolmogorov, expusă în capitolul I, care, după cum am afirmat, se bucură de un mare caracter de generalitate. Vom arăta cum se face această încadrare și preciziunile care decurg din ea pentru definirea legilor de probabilitate.

Cele două legi de probabilitate sînt definite pe intervale, considerînd că și punctul este un interval degenerat. Să vedem care este mulțimea de puncte de pe axa reală, care conține intervalele, și pe care pot fi definite legile de probabilitate, în conformitate cu teoria generală pe care am învățat-o.

Să considerăm evenimentul A_1, constînd în faptul că punctul ξ_1, luat la întîmplare, să cadă în interiorul intervalului (a_1, b_1), adică evenimentul $a_1 < \xi_1 < b_1$. În mod analog, fie evenimentul A_2, dat prin inegalitățile $a_2 < \xi_2 < b_2$. După teoria lui Kolmogorov, reuniunea a două evenimente A_1 și A_2 este de asemenea un eveniment. Prin urmare, $A_1 \cup A_2$ reprezintă de asemenea un eveniment și anume evenimentul ca un punct ξ luat la întîmplare să cadă sau în intervalul (a_1, b_1) sau în intervalul (a_2, b_2). Legile de probabilitate, pe care le folosim deci în statistică, vor trebui să fie definite și pe reuniuni de două intervale. Același raționament ne arată că vom trebui să luăm în considerare și reuniunile de un număr finit de intervale sau, chiar mai mult, o mulțime numărabilă de reuniuni de intervale.

Clasa de mulțimi de puncte de pe axa reală, obținută prin aplicarea operației de reuniune în număr finit sau de o infinitate numărabilă de ori asupra intervalelor, este mult mai cuprinzătoare decît clasa intervalelor.

Din teoria lui Kolmogorov rezultă însă că și intersecțiile de evenimente sînt tot evenimente. De asemenea diferența a două evenimente este tot un eveniment. De aceea, lor trebuie să le corespundă de asemenea o probabilitate.

Să însemnăm prin B clasa mulțimilor B, $B \in \mathrm{B}$ pentru care evenimentul $\xi \in B$ are un sens, adică admite o probabilitate corespunzătoare, îndeplinind condițiile generale expuse în capitolul I. Mulțimile B trebuie să se bucure de proprietatea că dacă luăm un număr finit sau o infinitate numărabilă de mulțimi B, reuniunea lor aparține de asemenea clasei B. De asemenea, aplicînd operația de intersecție la un număr finit sau la o infinitate numărabilă de mulțimi B, obținem tot o mulțime B. Complementara unei mulțimi B, este tot o mulțime B. Orice interval închis, deschis, semideschis, este o mulțime B.

Mulțimile B poartă numele de mulțimi boreliene, după numele matematicianului E. Borel *), iar B este clasa mulțimilor boreliene de pe axa reală.

Pentru orice mulțime boreliană B există prin urmare un număr $P(\xi \in B)$, care reprezintă probabilitatea ca evenimentul $\xi \in B$ să se realizeze. Numărul $P(\xi \in B)$ variază de la o lege de probabilitate la alta și caracterizează legea de probabilitate.

Fiind dată o lege de probabilitate, adică probabilitatea $P(\xi \in B)$, să considerăm funcția $F(x)$, definită prin relația

$$F(x) = P(\xi \leqslant x). \qquad \text{(VI,1.1)}$$

Funcția $F(x)$ există întotdeauna pentru că mulțimea punctelor ξ, îndeplinind condiția $\xi \leqslant x$, formează un interval, adică o mulțime boreliană. Ea poartă numele de *funcție de repartiție*.

Din faptul că $P(\xi \leqslant x)$ este o probabilitate, rezultă imediat că

$$0 \leqslant F(x) \leqslant 1,$$

oricare ar fi x.

În plus, dacă x crește necontenit, intervalul $\xi \leqslant x$ devine întreaga axă reală. Evenimentul $\xi \leqslant +\infty$ devine evenimentul sigur și în baza relației (VI.1.1) avem

$$F(+\infty) = 1.$$

Dimpotrivă, cînd x tinde către $-\infty$, evenimentul $\xi \geqslant x$ tinde către evenimentul imposibil și deci

$$F(-\infty) = 0.$$

Fie y un număr real mai mare decît x.

Intervalele $\xi \leqslant x$, $x < \xi \leqslant y$ nu au nici un punct comun. Pe de altă parte, reuniunea lor este intervalul $\xi \leqslant y$. Rezultă atunci din a patra axiomă din teoria lui Kolmogorov (Cap. I, punctul 3) că

$$P(\xi \leqslant y) = P(\xi \leqslant x) + P(x < \xi \leqslant y).$$

Ținînd seamă de (VI, 1.1) obținem

$$F(y) = F(x) + P(x < \xi \leqslant y), \quad (y > x),$$

sau

$$F(y) - F(x) = P(x < \xi \leqslant y), \quad (y > x). \qquad \text{(VI,1.2)}$$

Însă $P(x < \xi \leqslant y)$ reprezentînd probabilitatea evenimentului $x < \xi \leqslant y$ este un număr nenegativ.

Din relația precedentă rezultă astfel că dacă $y \geqslant x$,

$$F(y) - F(x) \geqslant 0.$$

$F(x)$ este deci o funcție nedescrescătoare de x.

Să punem în relația (VI, 1.2)

$$y = x + h.$$

Obținem

$$F(x + h) - F(x) = P(x < \xi \leqslant x + h), \; h > 0.$$

Dacă în această relație facem $h \to 0$, $F(x + h)$ tinde către o limită $F(x + 0)$, deoarece funcția

*) E. Borel (1871—1956), matematician francez, a publicat lucrări importante în domeniul analizei matematice, teoriei probabilităților și statistică. De asemenea a publicat cel dintîi studiu de strategia jocurilor, teorie importantă în matematica actuală, care cuprinde ca un caz particular programarea liniară și dinamică.

de repartiție $F(x)$ este o funcție nedescrescătoare. Pe de altă parte, evenimentul $x < \xi \leqslant x + h$, $h > 0$, cînd h tinde către zero, tinde către evenimentul imposibil. Prin urmare,

$$\lim_{\substack{h \to 0 \\ h > 0}} P(x < \xi \leqslant x + h) = 0$$

Din ultimele două relații rezultă

$$F(x) = F(x + 0), \tag{VI,1.3}$$

care ne arată că funcția de repartiție $F(x)$ este continuă la dreapta

Să scriem relația (VI, 1.2) sub forma echivalentă

$$F(x) - F(x - h) = P(x - h < \xi \leqslant x), \quad h > 0.$$

Dacă facem pe h să tindă către zero prin valori pozitive, funcția de repartiție $F(x - h)$, fiind nedescrescătoare, tinde către o limită, pe care o notăm prin $F(a - 0)$.

Evenimentul $x - h < \xi \leqslant x$ tinde către evenimentul $\xi = x$. Avem prin urmare:

$$\lim_{\substack{h \to 0 \\ h > 0}} P(x - h < \xi \leqslant x) = P(\xi = x)$$

Deducem din cele două relații de mai sus:

$$F(x) = F(x - 0) + P(\xi = x). \tag{VI,1.3'}$$

Egalitățile (VI,1.3) și (VI,1.3') sînt foarte importante. Prima dintre ele ne arată că valoarea funcției de repartiție $F(x)$, într-un punct x, este egală cu limita ei la dreapta. A doua exprimă faptul că valoarea funcției de repartiție $F(x)$ în punctul x este egală cu limita ei la stînga la care se adaugă probabilitatea obținerii evenimentului $\xi = x$. Dacă această ultimă probabilitate este egală cu zero, $F(x)$ este continuă și la stînga și deci este continuă în punctul x. În cazul contrar, cînd $P(\xi = x) > 0$, funcția de repartiție admite un punct de discontinuitate în punctul x.

Să observăm acum că în baza proprietăților pe care le-am demonstrat, relația (VI,1.2) poate fi extinsă la toate tipurile de intervale.

Avem, astfel:

$$\begin{aligned} P(x < \xi \leqslant y) &= F(y) - F(x), \\ P(x \leqslant \xi < y) &= F(y - 0) - F(x - 0), \\ P(x < \xi < y) &= F(y - 0) - F(x), \\ P(x \leqslant \xi \leqslant y) &= F(y) - F(x - 0), \end{aligned} \tag{VI,1.4}$$

pentru oricare $y > x$.

Probabilitatea ca punctul ξ să cadă într-un interval închis, deschis, semideschis, este astfel complet determinată cu ajutorul funcției de repartiție $F(x)$.

În mod invers, se poate arăta că orice funcție $F(x)$ nedescrescătoare, continuă la dreapta, îndeplinind condițiile

$$0 \leqslant F(x) \leqslant 1,$$

$$F(-\infty) = 0, \quad F(+\infty) = 1,$$

poate fi considerată ca funcția de repartiție a unei legi unice de probabilitate, a cărei probabilitate este definită prin egalitatea

$$P(\xi \leqslant x) = F(x).$$

Prin urmare, orice lege de probabilitate unidimensională poate fi complet definită prin funcția ei de repartiție.

În aceasta rezidă importanța cunoașterii funcției de repartiție.

Sîntem acum în măsură să definim cele două legi de probabilitate mai des întîlnite în statistică.

a) *Legi de probabilitate discrete.* În acest caz, există un număr finit de valori x_i cu probabilitățile respective p_i, $(i = 1, 2, \ldots, n)$, $p_i \geqslant 0$, $\sum_{i=1}^{n} p_i = 1$. Reprezentarea grafică a acestei distribuții a fost dată, pentru $i = 6$, în figura 1.

Funcția de repartiție se reprezintă ca o curbă cu trepte, avînd punctele de discontinuitate sau treptele în punctele x_i, $(i = 1, 2, \ldots, n)$. Aceasta rezultă clar din exemplul 26.

Punctele x_i pot fi și în număr infinit, formînd însă o mulțime numărabilă, cu condiția însă ca în orice interval finit să existe un număr finit de puncte x_i.

Astfel, legea Poisson, pe care o vom studia mai departe, ia valorile întregi $\alpha = \{0,1,2,\ldots\}$. Probabilitatea de a lua valoarea α este dată de expresia.

$$P_\alpha = \frac{\lambda^\alpha}{\alpha!} e^{-\lambda}.$$

Sîntem tot în cazul unei legi de probabilitate discrete deoarece în fiecare interval finit nu găsim decît un număr finit de puncte α.

b) *Legi de probabilitate continue.* În cazul cînd funcția de repartiție $F(x)$ admite în fiecare punct x o derivată, funcție continuă de x,

$$F'(x) = f(x)$$

legea de probabilitate este continuă.

Relațiile (VI, 1.4) ne dau:

$$P(x < \xi < x + \mathrm{d}x) = F(x + \mathrm{d}x) - F(x) = f(x)\,\mathrm{d}x$$

prin urmare probabilitatea de a obține o valoare cuprinsă între x și $x + \mathrm{d}x$. Funcția $f(x)$ poartă numele de *densitate de probabilitate.*

Deoarece $F(x)$ este nedescrescătoare, avem întotdeauna

$$f(x) \geqslant 0$$

ceea ce concordă și cu faptul că $f(x)\mathrm{d}x$ reprezintă o probabilitate.

Tot din relațiile (VI, 1.4) rezultă

$$P(\alpha < \xi < \beta) = \int_\alpha^\beta f(t)\mathrm{d}t.$$

Probabilitatea evenimentului $\alpha < \xi < \beta$ este prin urmare egală cu aria delimitată de axa Ox, curba $y = f(x)$ și dreptele $x = \alpha$, $x = \beta$.

Avem

$$\int_{-\infty}^{+\infty} f(t)\mathrm{d}t = 1,$$

sau dacă legea de probabilitate ia valori cuprinse numai în intervalul (a, b), ca în cazul figurii 3,

$$\int_a^b f(t)\mathrm{d}t = 1.$$

Scanned with OKEN Scanner

INTEGRALA STIELTJES. Să considerăm intervalul semideschis

$$I = (a < \xi \leqslant b),$$

unde a și b sînt două numere reale $(a < b)$.

Fie

$$a = x_0 < x_1 < \ldots < x_n = b$$

o diviziune a intervalului $I = (a, b)$ în n intervale

$$I_m = (x_{m-1} < \xi \leqslant x_m), \quad (m = 1,2,\ldots,n),$$

Fie ξ_m o valoare oarecare cuprinsă în intervalul I_m.

Se consideră suma

$$\sum_n = \sum_{m=1}^{n} \varphi(\xi_m)\,[F(x_m) - F(x_{m-1})], \tag{VI,1.5}$$

unde $\varphi(x)$ este o funcție definită pe intervalul (a, b), iar $F(x)$ o funcție de repartiție.

Dacă oricare ar fi diviziunea

$$I = I_1 + I_2 + \ldots + I_n,$$

Σ_n tinde către o limită ΣI cînd $n \to \infty$ și lungimea celui mai mare din intervalele I_m tinde către zero, vom nota această limită prin

$$\int_a^b \varphi(x)\mathrm{d}F(x).$$

Ea poartă numele de integrala lui Stieltjes a funcției $\varphi(x)$ în intervalul (a, b). Dacă acest proces de limită are loc, $\varphi(x)$ este o funcție integrabilă pe (a, b). Se poate arăta că o funcție continuă este integrabilă.

Integrala lui Stieltjes generalizează integrala lui Riemann, reducîndu-se la aceasta din urmă pentru $F(x) = x$.

Să considerăm cazul unei legi de probabilitate discrete:

$$\begin{pmatrix} a_1, & a_2, \ldots, & a_s \\ p_1, & p_2, \ldots, & p_s \end{pmatrix}. \tag{VI,1.6}$$

Există prin urmare numai s valori a_i, $(i = 1, 2,\ldots, s)$ care pot fi luate respectiv cu probabilitățile p_i, $(i = 1, 2,\ldots, s)$:

$$\sum_{i=1}^{s} p_i = 1.$$

Fie (a, b) un interval care conține punctele a_i, $(i = 1, 2,\ldots, s)$, $\varphi(x)$ fiind o funcție continuă în intervalul (a, b), ne propunem să calculăm integrala Stieltjes

$$\int_a^b \varphi(x)\mathrm{d}F(x),$$

unde $F(x)$ este funcția de repartiție a legii de probabilitate (VI,1.6).

În suma Σ_n dată de (VI,1.5) să notăm respectiv prin I_{i_1}, I_{i_2},..., I_{i_s} intervalele care conțin punctele $a_1, a_2,\ldots, a_s$ și să ne alegem punctele arbitrare $\xi_1, \xi_2,\ldots, \xi_s$ respectiv egale cu $a_1, a_2,\ldots, a_s$.

Funcția de repartiție $F(x)$, fiind o funcție cu trepte, avînd punctele de discontinuitate în a_i, $(i = 1, 2,..., s)$, pentru orice interval $I_m = (x_{m-1} < \xi \leqslant x_m)$, avem

$$F(x_m) - F(x_{m-1}) = 0.$$

Pentru intervalul

$$I_{j_k} = (x_{j_k-1} < \xi \leqslant x_{j_k}),$$

care conține punctul a_k, avem

$$F(x_{j_k}) - F(x_{j_k-1}) = p_k, \qquad (k = 1, 2,..., s),$$

deoarece funcția de repartiție $F(x)$ admite în punctul p_k un salt egal cu p_k.

Rezultă atunci că:

$$\sum_n = \sum_{m=1}^{n} \varphi(\xi_m)(F(x_m) - F(x_{m-1})) = \sum_{k=1}^{s} p_k \varphi(a_k).$$

Suma Σ_n rămîne mereu egală cu expresia:

$$\sum_{k-1}^{s} p_k \varphi(a_k),$$

care nu depinde de n, cînd facem pe $n \to \infty$ și cel mai mare interval I_m să tindă către zero, avînd grija să alegem punctele $\xi_{j_k} = a_k$, $(k = 1, 2,..., s)$.

Rezultă

$$\int_a^b \varphi(x)\mathrm{d}F(x) = \sum_{k=1}^{s} p_k \varphi(a_k). \qquad \text{(VI,1.7)}$$

Prin urmare, în cazul unei legi de probabilitate discrete, integrala Stieltjes se reduce la o sumă obișnuită. Expresia (VI,1.7) coincide cu valoarea medie a funcției $\varphi(x)$.

În cazul unei legi de probabilitate continue, derivabile, avem:

$$\mathrm{d}F(x) = f(x)\mathrm{d}x,$$

$$\int_a^b \varphi(x)\mathrm{d}F(x) = \int_a^b \varphi(x)f(x)\mathrm{d}x. \qquad \text{(VI,1.8)}$$

Integrala Stieltjes devine integrala definită obișnuită pe intervalul (a, b), adică integrala Riemann. Relația (VI, 1.8) ne dă valoarea medie a funcției $\varphi(x)$ pe intervalul (a, b).

Observăm că, atît pentru legea de probabilitate discretă, cît și pentru cea continuă, valoarea medie a funcției $\varphi(x)$ se exprimă prin integrala Stieltjes:

$$\int_a^b \varphi(x)\mathrm{d}F(x).$$

Această integrală exprimă valoarea medie a lui $\varphi(x)$ pentru orice lege de probabilitate definită prin funcția ei de repartiție $F(x)$. De aceea ea constituie instrumentul matematic, care se impune în mod natural pentru studiul valorilor medii ale legilor de probabilitate.

CÎTEVA PROPRIETĂȚI ALE INTEGRALEI STIELTJES. Cele mai importante din proprietățile integralei definite, expuse în cărțile clasice de calcul integral și diferențial, se păstrează și pentru integrala Stieltjes.

Scanned with OKEN Scanner

Avem astfel:

$$\int_a^b (\varphi_1(x) + \varphi_2(x))\, dF(x) = \int_a^b \varphi_1(x) dF(x) + \int_a^b \varphi_2(x) dF(x),$$

unde $\varphi_1(x)$ și $\varphi_2(x)$ sînt două funcții integrabile pe (a, b);

$$\int_a^b \varphi(x) dF(x) = \int_a^c \varphi(x) dF(x) + \int_c^b \varphi(x) dF(x),$$

cu condiția ca $\varphi(x)$ să fie integrabilă pe intervalele (a, b), (a, c), (c, b);

$$\int_a^b h\varphi(x) dF(x) = h \int_a^b \varphi(x) dF(x),$$

unde h este o constantă.

Notînd prin φ' și φ'' marginea inferioară și marginea superioară a funcției $\varphi(x)$ în intervalul (a, b), avem

$$\varphi'[F(b) - F(a)] \leqslant \int_a^b \varphi(x) dF(x) \leqslant \varphi''[F(b) - F(a)].$$

De asemenea putem scrie

$$\left| \int_a^b \varphi(x) dF(x) \right| \leqslant \int_a^b |\varphi(x)|\, dF(x).$$

Toate aceste proprietăți sînt utile la calculul valorilor medii.

OBSERVAȚIE IMPORTANTĂ. În capitolele IV și V am stabilit anumite proprietăți ale valorilor medii. Demonstrațiile le-am dat pentru cazul legilor de probabilitate discrete, utilizînd sumele finite cu care se exprimă valorile medii. Demonstrațiile nu se schimbă însă dacă exprimăm aceste sume finite cu ajutorul integralei Stieltjes și ținem seama de proprietățile ei mai sus arătate. Rezultă că proprietățile arătate sînt adevărate nu numai pentru legile de probabilitate discrete, ci pentru cele mai generale legi de probabilitate, definite prin funcția lor de repartiție $F(x)$.

În particular, ele sînt valabile și pentru legile de probabilitate continue.

EXEMPLUL 26. Se aruncă o monedă de trei ori. Apariția uneia din fețe este o variabilă aleatoare care are următoarea lege de probabilitate:

$$X \begin{bmatrix} 0, & 1, & 2, & 3 \\ \frac{1}{8}, & \frac{3}{8}, & \frac{3}{8}, & \frac{1}{8} \end{bmatrix}.$$

Ni se cere să determinăm valoarea funcției de repartiție pentru următoarele valori ale lui x:

a) $x = -2$;

b) $x = 2{,}5$;

c) $x = 3$.

Dacă notăm cu ξ orice valoare a variabilei noastre mai mică decît o valoare dată x, vom avea, prin definiție, funcția de repartiție $F(x)$, egală cu probabilitatea ca variabila ξ să ia o valoare mai mică sau egală cu x.

Deci,

$$F(x) = P(\xi \leqslant x).$$

a) În exemplul luat dacă $x = -2$, avem

$$F(-2) = P(\xi < -2) = 0.$$

Sîntem în cazul unui eveniment imposibil.

b) Dacă $x = 2{,}5$, funcția de repartiție în acest punct este

$$F(2{,}5) = P(\xi < 2{,}5).$$

Cînd poate avea loc un asemenea eveniment?

Este clar că un asemenea eveniment se verifică atunci cînd fața considerată a monedei a ieșit sau de zero ori, sau o singură dată sau de două ori (în cele trei aruncări). Întrucît probabilitățile respective sînt $\frac{1}{8}$, $\frac{3}{8}$ și $\frac{3}{8}$, funcția de repartiție este dată de suma lor.

Deci,

$$F(2{,}5) = P(\xi < 2{,}5) = \frac{1}{8} + \frac{3}{8} + \frac{3}{8} = \frac{7}{8}.$$

c) Dacă $x = 4$, funcția de repartiție în acest punct este

$$F(4) = P(\xi < 4).$$

Valoarea acestei funcții de repartiție este dată de suma probabilităților ca să se verifice — în cele trei aruncări ale monedei — apariția

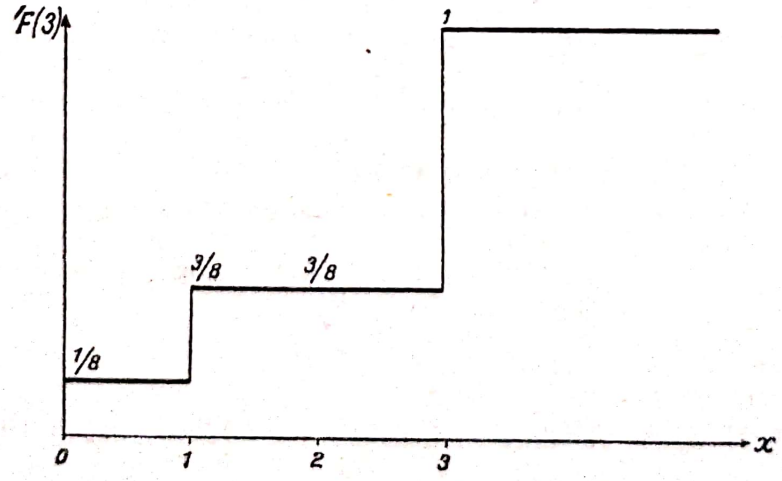


Fig. 4

feței care ne interesează prin una din cele patru valori posibile (sau 0, sau 1, sau 2, sau 3). Suma acestor probabilități este egală cu 1. Este

sigur că fața considerată va apare fie de trei ori, fie de două ori, fie o singură dată, fie niciodată; alte posibilități nu există.

Deci,

$$F(4) = P(\xi \leqslant 4) = \frac{1}{8} + \frac{3}{8} + \frac{3}{8} + \frac{1}{8} = 1.$$

Funcția de repartiție este reprezentată grafic ca în figura 4.

Funcția de repartiție, reprezentată grafic, se prezintă ca o funcție cu trepte. În fiecare din punctele 0, 1, 2, 3 ea admite puncte de discontinuitate.

2. Funcție caracteristică

Să considerăm variabila aleatoare ξ. Valoarea medie a expresiei $e^{it\xi}$, unde t este o variabilă reală iar i unitatea imaginară, poartă numele de funcție caracteristică.

În baza acestei definiții, dacă variabila ξ poate lua valorile x_j, $(j = 1, 2, \ldots, n)$, cu probabilitățile respective p_j, $(j = 1, 2, \ldots, n)$, funcția caracteristică corespunzătoare este

$$\varphi(t) = \sum_{j=1}^{n} p_j e^{itx_j}. \qquad \text{(VI,2.1)}$$

În cazul unei variabile continue ξ, cu densitatea de probabilitate $f(x)$, funcția caracteristică este

$$\varphi(t) = \int_{-\infty}^{+\infty} e^{itx} f(x)\,dx. \qquad \text{(VI, 2.2)}$$

Ambele formule sînt cuprinse în formula mai generală.

$$\varphi(t) = \int_{-\infty}^{+\infty} e^{itx}\,dF(x).$$

Două legi de probabilitate care au aceeași funcție caracteristică sînt identice. De aici provine și numele dat acestei funcții.

Se știe că

$$e^{itx} = 1 + \frac{itx}{1!} + \frac{(itx)^2}{2!} + \ldots + \frac{(itx)^n}{n!} + \ldots$$

Dacă înlocuim această desfășurare în serie a lui e^{itx} în relația (VI,2.2), se poate demonstra că putem scrie

$$\varphi(t) = \sum_{n=0}^{\infty} \frac{i^n t^n}{n!} \int_{-\infty}^{+\infty} x^n f(x)\,dx.$$

Însă

$$\int_{-\infty}^{+\infty} x^n f(x)\,dx = M_n$$

reprezintă momentul M_n, de ordinul n, al lui ξ, cu condiția ca acest moment să existe.

Comparînd relațiile precedente, rezultă

$$\varphi(t) = \sum_{n=0}^{\infty} \frac{i^n t^n}{n!} M_n. \qquad \text{(VI,2.3)}$$

Obținem astfel o exprimare simplă a funcției caracteristice în funcție de momentele variabilei ξ. Această relație prezintă interes, pentru că în anumite împrejurări calculul funcției caracteristice se poate face mai ușor decît al momentelor. Într-adevăr, derivînd de n ori ambii membri ai relației (VI,2.3) și făcînd $t = 0$, obținem

$$M_n = \frac{1}{i^n}\left[\frac{\mathrm{d}^n \varphi(t)}{\mathrm{d}t^n}\right]_{t=0}. \qquad \text{(VI,2.4)}$$

Deci derivata de ordinul n a funcției caracteristice în punctul $t = 0$, împărțită la i^n, ne dă momentul de ordinul n.

Vom pune acum în evidență o proprietate însemnată a funcției caracteristice.

Să considerăm astfel două variabile aleatoare ξ_1 și ξ_2, independente între ele, cu funcțiile caracteristice respective $\varphi_1(t)$ și $\varphi_2(t)$. Însemnînd prin $\xi = \xi_1 + \xi_2$, suma variabilelor ξ_1 și ξ_2 și prin $\varphi(t)$ funcția caracteristică a variabilei ξ, se poate ușor demonstra că

$$\varphi(t) = \varphi_1(t)\varphi_2(t). \qquad \text{(VI,2.5)}$$

Într-adevăr, prin definiție

$$\varphi_1(t) = M(e^{i\xi_1 t}), \quad \varphi_2(t) = M(e^{i\xi_2 t}),$$

$$\varphi(t) = M(e^{i\xi t}) = M(e^{i(\xi_1+\xi_2)t}) = M(e^{i\xi_1 t}\, e^{i\xi_2 t}) = M(e^{i\xi_1 t})\, M(e^{i\xi_2 t}) = \varphi_1(t)\varphi_2(t),$$

deoarece variabilele $e^{i\xi_1 t}$ și $e^{i\xi_2 t}$ sînt independente între ele, iar valoarea medie a unui produs de variabile independente este egală cu produsul valorilor medii ale celor două variabile.

În același mod se demonstrează relația mai generală

$$\varphi(t) = \varphi_1(t)\varphi_2(t)\ldots\varphi_n(t), \qquad \text{(VI,2.6)}$$

unde $\varphi(t)$ înseamnă funcția caracteristică a variabilei $\xi_1+\xi_2+\ldots+\xi_n$, ξ_1, $\xi_2, \ldots, \xi_n$ fiind n variabile independente, iar $\varphi_i(t)$ funcția caracteristică a variabilei ξ_i, $(i = 1,2,\ldots, n)$.

De mare importanță este și următoarea teoremă:

Fiind dat un șir de variabile aleatoare cu funcțiile de repartiție $F_1(x)$, $F_2(x)$,... și cu funcțiile caracteristice $\varphi_1(t)$, $\varphi_2(t)$,..., o condiție necesară și suficientă pentru ca șirul $F_i(x)$, $(i=1,2,\ldots)$, să tindă către o funcție de repartiție $F(x)$ a unei variabile aleatoare este ca șirul funcțiilor caracteristice $\varphi_i(t)$, $(i = 1,2,\ldots, n\ldots)$, să tindă către o limită $\varphi(t)$, continuă în punctul $t = 0$. Această funcție $\varphi(t)$ este funcția caracteristică a funcției de repartiție limită $F(x)$.

Cu alte cuvinte dacă avem

$$\lim_{n\to\infty} \varphi_n(t) = \varphi(t)$$

rezultă

$$\lim_{n\to\infty} F_n(x) = F(x).$$

Vom avea ocazia să aplicăm de mai multe ori în cele ce urmează această proprietate însemnată a funcției caracteristice.

Scanned with OKEN Scanner

Capitolul VII

Legi de probabilitate clasice

1. Legea binomială

La capitolul III, cînd ne-am ocupat de schema clasică a urnei Bernoulli cu două stări, am dedus o formulă generală (III, 1.2) care ne-a permis să calculăm probabilitatea ca în n extracții să se verifice de α ori apariția unei bile de o anumită culoare.

Teoretic, probabilitățile apariției une bile de o anumită culoare de α ori ($\alpha = 0,1,2, \ldots, n$), în n extracții, sînt date — după cum am văzut — de termenii succesivi ai binomului lui Newton, fiecare termen fiind expresia formulei generale $C_n^\alpha p^\alpha q^{n-\alpha}$.

Astfel, pentru cazul cînd facem, trei extracții, bila albă poate să ne apară de trei ori, de două ori, o singură dată sau niciodată (apărînd de trei ori bila neagră), cu probabilitățile respective:

$$C_3^3 p^3 q^0 = p^3;\ C_3^2 p^2 q^{3-2} = 3p^2 q;\ C_3^1 p^1 q^{3-1} = 3pq^2;\ C_3^0 p^0 q^{3-0} = q^3.$$

Întrucît evenimentele sînt incompatibile (de exemplu, faptul că bila albă a ieșit de trei ori, exclude celelalte trei posibilități), probabilitatea evenimentului reunit, este egală cu suma probabilităților fiecărui eveniment în parte, adică:

$$p^3 + 3p^2 q + 3pq^2 + q^3 = 1.$$

Observăm că într-adevăr legea de probabilitate pentru urna lui Bernoulli este identică cu modelul matematic al desfășurării binomului lui Newton. Tocmai datorită acestui fapt, legea de probabilitate de forma

$$X\begin{bmatrix} 0, & 1, & 2, & 3, & \ldots, & n \\ C_n^0 p^0 q^{n-0}, & C_n^1 p^1 q^{n-1}, & C_n^2 p^2 q^{n-2}, & C_n^3 p^3 q^{n-3}, & \ldots, & C_n^n p^n q^0 \end{bmatrix}$$

se numește *lege binomială*. Formula

$$P_n(\alpha) = C_n^\alpha p^\alpha q^{n-\alpha}$$

o sintetizează.

Legea binomială are anumite trăsături caracteristice pe care le vom releva, pe scurt, în cele ce urmează:

În primul rînd, legea binomială este o lege de tip discret, discontinuă. În reprezentarea grafică acest lucru iese clar în evidență (vezi fig. 5).

EXEMPLUL 27. Se aruncă o monedă de 12 ori. Să se reprezinte grafic legea de probabilitate a numărului de cazuri în care ne poate apare una din fețele monedei.

Este evident că numărul de cazuri α, teoretic posibile, în care ne poate apare una din fețe, variază de la 0 la 12. Legea de probabilitate este ilustrată în tabelul 2.

Tabelul 2

$$P_{12(\alpha)} = C_{12}^{\alpha}\left(\frac{1}{2}\right)^{\alpha} \cdot \left(\frac{1}{2}\right)^{12-\alpha}$$

De cîte ori ne apare fața care ne interesează (α)	Probabilitatea $P_{12(\alpha)}$	Probabilitatea cumulată $\Sigma P_{12(\alpha)} = F(\alpha)$
0	$C_{12}^{0}\left(\frac{1}{2}\right)^{0} \cdot \left(\frac{1}{2}\right)^{12} = \frac{1}{4\,096}$	$\frac{1}{4\,096}$
1	$C_{12}^{1}\left(\frac{1}{2}\right)^{1} \cdot \left(\frac{1}{2}\right)^{11} = \frac{12}{4\,096}$	$\frac{13}{4\,096}$
2	$C_{12}^{2}\left(\frac{1}{2}\right)^{2} \cdot \left(\frac{1}{2}\right)^{10} = \frac{66}{4\,096}$	$\frac{79}{4\,096}$
3	$C_{12}^{3}\left(\frac{1}{2}\right)^{3} \cdot \left(\frac{1}{2}\right)^{9} = \frac{220}{4\,096}$	$\frac{299}{4\,096}$
4	$C_{12}^{4}\left(\frac{1}{2}\right)^{4} \cdot \left(\frac{1}{2}\right)^{8} = \frac{495}{4\,096}$	$\frac{794}{4\,096}$
5	$C_{12}^{5}\left(\frac{1}{2}\right)^{5} \cdot \left(\frac{1}{2}\right)^{7} = \frac{792}{4\,096}$	$\frac{1\,586}{4\,096}$
6	$C_{12}^{6}\left(\frac{1}{2}\right)^{6} \cdot \left(\frac{1}{2}\right)^{6} = \frac{924}{4\,096}$	$\frac{2\,510}{4\,096}$
7	$C_{12}^{7}\left(\frac{1}{2}\right)^{7} \cdot \left(\frac{1}{2}\right)^{5} = \frac{792}{4\,096}$	$\frac{3\,302}{4\,096}$
8	$C_{12}^{8}\left(\frac{1}{2}\right)^{8} \cdot \left(\frac{1}{2}\right)^{4} = \frac{495}{4\,096}$	$\frac{3\,797}{4\,096}$
9	$C_{12}^{9}\left(\frac{1}{2}\right)^{9} \cdot \left(\frac{1}{2}\right)^{3} = \frac{220}{4\,096}$	$\frac{4\,017}{4\,096}$
10	$C_{12}^{10}\left(\frac{1}{2}\right)^{10} \cdot \left(\frac{1}{2}\right)^{2} = \frac{66}{4\,096}$	$\frac{4\,083}{4\,096}$
11	$C_{12}^{11}\left(\frac{1}{2}\right)^{11} \cdot \left(\frac{1}{2}\right)^{1} = \frac{12}{4\,096}$	$\frac{4\,095}{4\,096}$
12	$C_{12}^{12}\left(\frac{1}{2}\right)^{12} \cdot \left(\frac{1}{2}\right)^{0} = \frac{1}{4\,096}$	$\frac{4\,096}{4\,096} = 1$

Din figura 5 se observă că probabilitățile valorilor $\alpha = 0,1,2, \ldots$ cresc pînă la un anumit punct după care încep să descrească, dispunîndu-se simetric în jurul probabilității corespunzătoare valorii medii $np = 6$.

Această comportare a probabilității $P_{n(\alpha)}$, (întîi creștere și apoi descreștere, în funcție de creșterea numărului α), observată în exemplul de

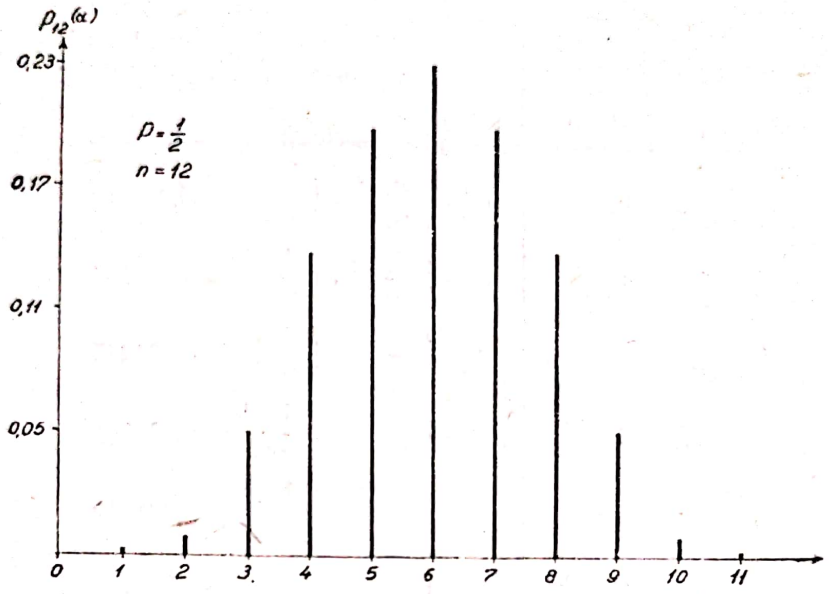


Fig. 5

mai sus, este o trăsătură caracteristică generală pentru legile binomiale de probabilitate, în care evenimentul elementar se verifică cu probabilitatea $p = \frac{1}{2}$.

Legile binomiale în care probabilitatea evenimentului elementar este mai mică sau mai mare decît $\frac{1}{2}$, sînt asimetrice, cînd n este mic; însă, cînd n crește necontenit, asimetria dispare treptat, legea binomială avînd tendința de a se apropia de *legea normală* de probabilitate, pe care o tratăm mai jos.

EXEMPLUL 28. În urma unor observații îndelungate s-a constatat că la întreprinderea industrială A, probabilitatea ca consumul de energie electrică să fie normal (să nu depășească un număr determinat de kWh în 24 ore) este egală cu $\frac{9}{10}$. Ni se cere să aflăm probabilitatea ca consumul de energie electrică să fie normal:

a) în 6 zile consecutive, timp de 0,1,2, ..., 6 zile și
b) în 12 zile consecutive, timp de 0, 1, 2, ..., 12 zile.

Să se reprezinte grafic legile de probabilitate respective.

În tabelele 3 și 4 sînt determinate probabilitățile respective $P_{6(\alpha)}$ și $P_{12(\alpha)}$.

Tabelul 3

$$P_{6(\alpha)} = C_6^{\alpha}\left(\frac{9}{10}\right)^{\alpha} \cdot \left(\frac{1}{10}\right)^{6-\alpha}$$

Numărul zilelor cu consum normal (α)	Probabilitățile
0	$C_6^0 \cdot \left(\frac{9}{10}\right)^0 \cdot \left(\frac{1}{10}\right)^6 = 0{,}000001$
1	$C_6^1 \cdot \left(\frac{9}{10}\right)^1 \cdot \left(\frac{1}{10}\right)^5 = 0{,}000054$
2	$C_6^2 \cdot \left(\frac{9}{10}\right)^2 \cdot \left(\frac{1}{10}\right)^4 = 0{,}001215$
3	$C_6^3 \cdot \left(\frac{9}{10}\right)^3 \cdot \left(\frac{1}{10}\right)^3 = 0{,}024580$
4	$C_6^4 \cdot \left(\frac{9}{10}\right)^4 \cdot \left(\frac{1}{10}\right)^2 = 0{,}098415$
5	$C_6^5 \cdot \left(\frac{9}{10}\right)^5 \cdot \left(\frac{1}{10}\right)^1 = 0{,}354294$
6	$C_6^6 \cdot \left(\frac{9}{10}\right)^6 \cdot \left(\frac{1}{10}\right)^0 = 0{,}531441$
	$\Sigma C_6^{\alpha} \cdot \left(\frac{9}{10}\right)^{\alpha} \cdot \left(\frac{1}{10}\right)^{6-\alpha} = 1$

Tabelul 4

$$P_{12(\alpha)} = C_{12}^{\alpha} \cdot \left(\frac{9}{10}\right)^{\alpha} \cdot \left(\frac{1}{10}\right)^{12-\alpha}$$

Numărul zilelor de consum normal (α)	Probabilitățile
0	$C_{12}^0 \cdot \left(\frac{9}{10}\right)^0 \cdot \left(\frac{1}{10}\right)^{12} = 0{,}000000000001$
1	$C_{12}^1 \cdot \left(\frac{9}{10}\right)^1 \cdot \left(\frac{1}{10}\right)^{11} = 0{,}000000000108$
2	$C_{12}^2 \cdot \left(\frac{9}{10}\right)^2 \cdot \left(\frac{1}{10}\right)^{10} = 0{,}000000005341$
3	$C_{12}^3 \cdot \left(\frac{9}{10}\right)^3 \cdot \left(\frac{1}{10}\right)^9 = 0{,}000000160380$
4	$C_{12}^4 \cdot \left(\frac{9}{10}\right)^4 \cdot \left(\frac{1}{10}\right)^8 = 0{,}000003247595$

Scanned with OKEN Scanner

(Continuare tabela 4)

$$P_{12(\alpha)} = C_{12}^{\alpha} \cdot \left(\frac{9}{10}\right)^{\alpha} \cdot \left(\frac{1}{10}\right)^{12-\alpha}$$

Numărul zilelor de consum normal (α)	Probabilitățile
5	$C_{12}^{5} \cdot \left(\frac{9}{10}\right)^{5} \cdot \left(\frac{1}{10}\right)^{7} = 0{,}000046737108$
6	$C_{12}^{6} \cdot \left(\frac{9}{10}\right)^{6} \cdot \left(\frac{1}{10}\right)^{6} = 0{,}000491051484$
7	$C_{12}^{7} \cdot \left(\frac{9}{10}\right)^{7} \cdot \left(\frac{1}{10}\right)^{5} = 0{,}003788110648$
8	$C_{12}^{8} \cdot \left(\frac{9}{10}\right)^{8} \cdot \left(\frac{1}{10}\right)^{4} = 0{,}021308126795$
9	$C_{12}^{9} \cdot \left(\frac{9}{10}\right)^{9} \cdot \left(\frac{1}{10}\right)^{3} = 0{,}085232507185$
10	$C_{12}^{10} \cdot \left(\frac{9}{10}\right)^{10} \cdot \left(\frac{1}{10}\right)^{2} = 0{,}230127875386$
11	$C_{12}^{11} \cdot \left(\frac{9}{10}\right)^{11} \cdot \left(\frac{1}{10}\right)^{1} = 0{,}376572490268$
12	$C_{12}^{12} \cdot \left(\frac{9}{10}\right)^{12} \cdot \left(\frac{1}{10}\right)^{0} = 0{,}282429687701$
	$\Sigma C_{12}^{\alpha} \cdot \left(\frac{9}{10}\right)^{\alpha} \cdot \left(\frac{1}{10}\right)^{12-\alpha} = 1$

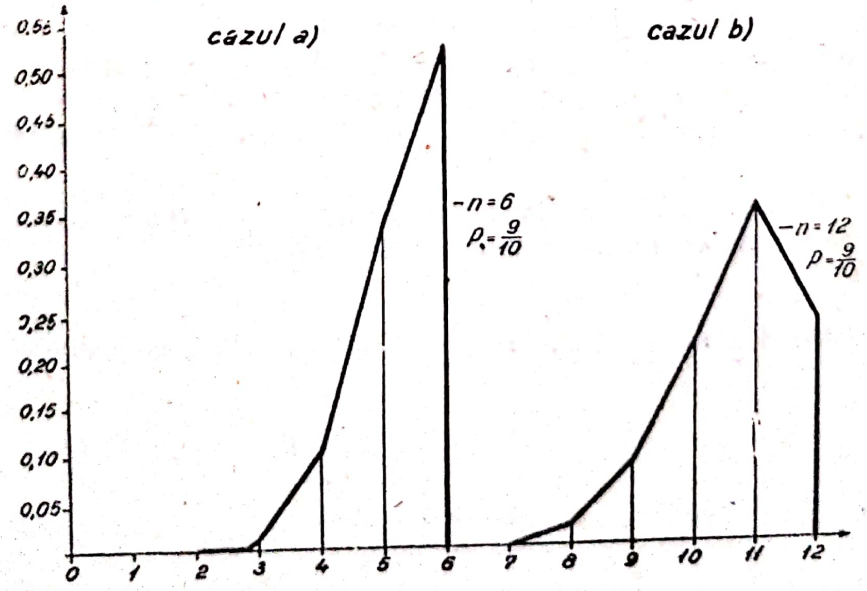


Fig. 6

Din datele tabelului 3 și din figura 6 se poate observa în decursul unei săptămîni (cazul *a*) evenimentul cel mai probabil este *să nu avem nici o depășire* a consumului de energie electrică; mai precis, în peste 53% de cazuri vom avea consum normal de energie electrică, timp de șase zile consecutive. Cu o probabilitate ceva mai mică (în 35% de cazuri), putem să ne așteptăm ca în una din cele șase zile să avem un consum supranormal, iar cu o probabilitate și mai mică (în aproximativ 9% de cazuri), putem să ne așteptăm ca în două din zilele săptămînii să avem un consum supranormal. În sfîrșit, cu o probabilitate, practic egală cu 0 (eveniment imposibil), putem să ne așteptăm ca în majoritatea zilelor săptămînii sau în fiecare zi să avem o depășire a consumului normal:

$$P_{6(0)} + P_{6(1)} + P_{6(2)} + P_{6(3)} \approx 0.$$

Mai observăm apoi că în cazul *a*), diagrama legii de probabilitate este vădit asimetrică. Ea crește pînă la un punct foarte apropiat de media variabilei aleatoare $np = 5{,}4$.

O altă comportare are legea de probabilitate pentru $n = 12$ (cazul *b*), p fiind același ca și în cazul *a*). Observăm că probabilitatea crește odată cu creșterea variabilei α și în punctul 11 al acesteia (foarte apropiat de $np = 10{,}8$), începe să descrească însă nu chiar atît de asimetric ca în cazul *a*).

Interesant de urmărit este faptul că în decursul celor 12 zile, cel mai probabil este evenimentul ca timp de 11 zile să avem consum normal (în aproximativ 38% de cazuri); ceva mai mică este probabilitatea ca să avem consum normal în toate cele 12 zile (28%) și probabilitatea și mai mică (23%) este să avem în două zile o depășire a consumului de energie electrică.

Un eveniment, practic imposibil, este acela ca să avem depășiri de consum de energie electrică timp de peste 5 zile.

După cum se vede, în ambele cazuri *a*) și *b*), a apărut o valoare a numărului de zile cu consum normal de energie electrică, care este *cea mai probabilă*. Acestei valori i se spune — după cum știm — *valoare modală* sau *modul*.

Dacă o notăm cu α_0 pentru legea de probabilitate binomială, putem enunța următoarea teoremă cu privire la modul:

TEOREMĂ. *Valoarea cea mai probabilă* α_0 *satisface inegalitățile*

$$np - q \leqslant \alpha_0 \leqslant np + p. \qquad \text{(VII, 1.1)}$$

Într-adevăr, am văzut că legea de probabilitate binomială este dată de expresia

$$C_n^x p^x q^{n-x}, \qquad (x = 0, 1, 2, \ldots, n). \qquad \text{(VII, 1.2)}$$

Presupunînd că expresiile (VII,1.2) sînt aranjate în ordinea crescătoare a puterilor lui p, atunci raportul uneia dintre ele față de precedenta este

$$\frac{P_{n(\alpha+1)}}{P_{n(\alpha)}} = \frac{C_n^{\alpha+1} p^{\alpha+1} q^{n-\alpha-1}}{C_n^{\alpha} p^{\alpha} q^{n-\alpha}} = \frac{n!\, p^{\alpha+1}\, q^{n-\alpha-1}}{(\alpha+1)!\,(n-\alpha-1)!} \cdot \frac{\alpha!\,(n-\alpha)!}{n!\, p^{\alpha} q^{n-\alpha}} = \frac{n-\alpha}{\alpha+1} \cdot \frac{p}{q}. \qquad \text{(VII, 1.3)}$$

Pentru a determina valoarea lui α_0 trebuie să stabilim mai întîi care din următoarele trei relații este adevărată:

$$\frac{n-\alpha}{\alpha+1} \cdot \frac{p}{q} > 1, \quad \frac{n-\alpha}{\alpha+1} \cdot \frac{p}{q} = 1, \quad \frac{n-\alpha}{\alpha+1} \frac{p}{q} < 1. \qquad \text{(VII, 1.4)}$$

Dacă, de pildă, vrem să știm valorile lui α pentru care are loc inegalitatea $P_{n(\alpha+1)} > P_{n(\alpha)}$, trebuie să determinăm valorile lui α pentru care este valabilă inegalitatea

$$\frac{n-\alpha}{\alpha+1} \cdot \frac{p}{q} > 1$$

sau

$$(n-\alpha)p > (\alpha+1)q$$

de unde

$$np-q > \alpha.$$

Aceasta înseamnă că atîta timp cît valoarea lui α nu a depășit valoarea lui $np-q$, probabilitatea $P_{n(\alpha)}$ crește; această probabilitate descrește dacă avem $np-p < \alpha$. Dacă avem $\alpha = np-q$, atunci $P_{n(\alpha+1)} = P_{n(\alpha)}$.

Prin urmare rezultă că

$P_{n(\alpha+1)} = P_{n(\alpha)}$, dacă avem $\alpha = np-q$,

$P_{n(\alpha+1)} > P_{n(\alpha)}$, dacă avem $\alpha < np-q$,

$P_{n(\alpha+1)} < P_{n(\alpha)}$, dacă avem $\alpha > np-q$.

Valoarea cea mai probabilă α_0 a numărului α trebuie, deci, să satisfacă dubla inegalitate

$$np-q \leqslant \alpha_0 \leqslant np+p. \qquad \text{(VII, 1.5)}$$

Modulul α_0 poate fi determinat în mod unic, dar nu este exclus să avem două valori modale și aceasta se întîmplă cînd mărimea $np-q$ este un număr întreg. Cele două valori diferă între ele printr-o unitate și au probabilitățile egale.

Dacă ne referim la exemplul 28 observăm că (în cazul b) modulul α_0 are valoarea 11. Se verifică astfel inegalitatea (VII,1.5):

$$\frac{12 \cdot 9}{10} - \frac{1}{9} < 11 < \frac{12 \cdot 9}{10} + \frac{9}{10}$$

sau

$$10{,}7 < 11 < 11{,}7.$$

După cum se vede, pentru legea de probabilitate binomială, *valoarea-modul* diferă foarte puțin de *valoarea medie* np; în cazul b) din exemplul considerat avem $np = 10{,}8$.

Funcția de repartiție a legii binomiale $F(x)$ este definită de relația

$$F(x) = P(\alpha \leqslant x) = \sum_{\alpha=0}^{[x]} P_{n(\alpha)}. \qquad \text{(VII,1.6)}$$

unde am notat prin $[x]$ cel mai mare întreg conținut în x. Se vede că $F(x)$ a legii binomiale corespunde probabilităților cumulate $\sum_{\alpha=0}^{[x]} P_{n(\alpha)}$ (vezi tabelul 2) și este o funcție cu trepte admițînd discontinuități în fiecare dintre punctele 0, 1, 2, ..., $[x]$.

Graficul funcției de repartiție va fi o funcție nedescrescătoare cu trepte, în număr egal cu numărul punctelor de discontinuitate.

Funcția caracteristică a legii binomiale, este, prin definiție, următoarea :

$$\varphi(t) = M\,(e^{\alpha ti}) = \sum_{\alpha=0}^{n} e^{\alpha ti}\, C_n^{\alpha} p^{\alpha}\, q^{n-\alpha} = \sum C_n^{\alpha} (pe^{ti})^{\alpha} q^{n-\alpha} = \\ = (pe^{ti} + q)^n. \qquad \text{(VII,1.7)}$$

Folosind teoremele cu privire la valorile medii ale sumei și produsului de variabile independente, putem deduce formulele de calcul ale diferitelor momente *obișnuite* și *centrate* pentru legile binomiale.

Se ajunge astfel la următoarele egalități pentru primele patru momente obișnuite :

$$M(X) = np,$$

$$M_2(X) = n^2p^2 + npq,$$

$$M_3(X) = n^3p^3 + 3n^2p^2q + npq\,(1-2p), \qquad \text{(VII,1.8)}$$

$$M_4(X) = n^4p^4 + 6n^3p^3q + n^2p^2(7 - 18p + 11p^2) + \\ + np(1 - 7p + 12p^2 - 6p^3). \qquad \text{(VII,1.9)}$$

Pentru momentul obișnuit de ordinul r avem

$$M_r(X) = n^rp^r + A_1n^{r-1} + A_2\,n^{r-2} + \ldots + A_{n-1}\,n$$

adică un polinom de gradul r în raport cu n.

Cu ajutorul momentelor obișnuite putem calcula și *momentele centrate* $\mu_1, \mu_2, \ldots, \mu_r$, folosind relația

$$\mu_r = M_r - C_r^1 M_{r-1} M + C_r^2 M_{r-2} M^2 + \ldots + (-1)^{r-1} C_r^r M^r$$

obținem următoarele egalități pentru primele patru momente centrate :

$$\mu_1 = 0,$$

$$\mu_2 = M_2 - M^2 = npq,$$

$$\mu_3 = M_3 - 3M_2M + 2M^3 = npq(q - p), \qquad \text{(VII,1.10)}$$

$$\mu_4 = M_4 - 4M_3M + 6M_2M^2 - 3M^4 = 3n^2p^2q^2 + npq(1 - 6pq). \qquad \text{(VII,1.11)}$$

Legea binomială este una dintre cele mai vechi din calculul probabilităților. J. Bernoulli, primul care a remarcat-o, a descoperit cu ajutorul ei legea numerelor mari (vezi Cap. XVIII).

2. Legea normală

Cercetările teoretice în domeniul calculului probabilităților, încă de acum 200 de ani, au arătat că într-un număr de cazuri din viața reală, întîlnim variabile aleatoare cu caracter de „erori accidentale" și care au legi de probabilitate de un anumit tip. S-a observat astfel, de exemplu, că rezultatele măsurării unei mărimi oarecare, a cărei valoare exactă este x, nu coincid cu această valoare ci sînt mai mari sau mai mici decît ea; abaterile în sens pozitiv sau negativ de la această valoare sînt tocmai acele „erori" accidentale la care ne-am referit mai sus. În general, rezultatul unor măsurări sau al unor experiențe este cu atît mai variat cu cît se cere o precizie mai mare. Erorile de observație sînt inerente măsurărilor de mare precizie și de aceea studiul lor prezintă un interes deosebit. Dacă, de pildă, cîntărim o anumită greutate de mai multe ori și dorim o precizie pînă la un miligram sau o fracțiune de miligram, rezultatele vor diferi mult între ele.

Dacă notăm cu x_0 rezultatul unei măsurări, atunci diferența

$$\xi = x - x_0$$

este tocmai eroarea de măsurare (observație).

Erorile de observație pot fi de două feluri:

— *erori sistematice*;

— *erori accidentale*.

În general, prin erori sistematice se înțeleg erori datorite imperfecțiunii aparatelor de măsurat; rezultatele vor fi influențate sistematic, în plus sau în minus, după construcția greșită a aparatului de măsurat. Aceste erori pot fi corectate.

Nu același lucru se poate spune despre erorile accidentale. Asemenea erori apar datorită unui complex de împrejurări care nu pot fi sesizate individual, nici în timpul procesului de măsurare nici după ce acesta a avut loc. Fiecare din împrejurările menționate, determină cîte o *eroare parțială*. Erorile parțiale sînt, în fond, niște variabile aleatoare, așa încît, *acțiunea fiecăreia dintre ele reprezintă numai o parte mică a acțiunii lor globale*, iar eroarea definitivă ξ care ne interesează de obicei este o rezultantă a tuturor erorilor accidentale parțiale.

De pildă, o eroare efectivă de măsură poate fi un rezultat al influenței mai multor cauze, fiecare din ele contribuind la eroarea de măsură înregistrată (presupunem că erorile sistematice au fost complet eliminate). Din aceste cauze generatoare de erori parțiale fac parte cele provenite din starea aparatului de măsurat influențat întîmplător, de factori atmosferici,

termici, mecanici etc. sau cauze subiective ale observatorului (particularități ale văzului sau alte particularități, și care la rîndul lor se schimbă în funcție de starea psihică sau fizică a observatorului).

În teoria probabilităților se demonstrează (vezi raționamentul teoretic în paragraful următor, cu privire la *teorema centrală a calculului probabilităților*), că *legea de probabilitate* a unei variabile aleatoare X, care este suma unui mare număr de variabile aleatoare $x_1, x_2, \ldots, x_n$, reciproc independente *) trebuie să fie apropiată de o lege de un tip determinat, *tipul legii normale*, oricare ar fi natura termenilor, *cu două condiții esențiale și anume: ca numărul n al cauzelor ce acționează să fie foarte mare și ca acțiunea fiecăreia dintre aceste cauze, față de acțiunea lor globală, să fie mică.* Formularea matematică precisă a acestei de-a doua condiții va fi făcută în paragraful următor.

Această teoremă domină întreaga teorie a erorilor. K. F. Gauss **) și P. S. Laplace ***) precum și alți cercetători au observat că erorile accidentale se repartizează *simetric* în jurul unei *valori medii*, după regulile date de așa numita *lege normală de probabilitate*. Studiul repartiției erorilor a dus (în a doua jumătate a secolului XIX-lea) la relevarea faptului că legea normală de probabilitate poate fi luată drept *model teoretic* pentru cercetarea probabilistică a aproape tuturor fenomenelor naturii. În prezent însă, legii normale nu i se mai atribuie un caracter general, recunoscîndu-se existența în natură și a multor fenomene care se apropie de modelul teoretic al altor legi de probabilitate.

Importanța legii normale de probabilitate a fost pusă în evidență de marii matematicieni Laplace și Gauss. Pentru acest motiv, este cunoscută sub numele de legea de probabilitate a lui Gauss sau de legea de probabilitate a lui Laplace. Meritul priorității revine însă lui Laplace și de aceea considerăm că este mai îndreptățit să se atribuie acestei legi numele de legea normală a lui Laplace.

---

*) Cercetări recente au dus la concluzia că independența variabilelor aleatoare care se însumează nu este o condiție esențială. Astfel, S. Bernștein a arătat că, cunoscuta teoremă a lui Liapunov se menține în condiții destul de puțin restrictive și pentru variabile dependente.

**) K. F. Gauss (1777—1855), unul din cei mai mari matematicieni ai lumii, a publicat lucrări de mare importanță în domeniul teoriei numerelor, analizei matematice, geometriei, astronomiei, geodeziei etc.

A publicat în anul 1801 lucrarea sa fundamentală *Disquisitiones aritmeticae*, în care a pus bazele teoriei moderne a numerelor. În geometrie a făcut primele cercetări importante în domeniul curburii suprafețelor. A descoperit metoda celor mai mici pătrate. A fost preocupat de bazele geometriei, găsindu-se după moartea lui însemnări importante în legătură cu geometria neeuclidiană.

***) P. S. Laplace (1749—1827), ilustru matematician, astronom și fizician francez, a publicat tratatul *Teoria analitică a probabilităților* (1813), care reprezintă nu numai cea dintîi expunere de ansamblu a calculului probabilităților ca o disciplină matematică de sine stătătoare, ci și o sursă inepuizabilă de idei pentru dezvoltarea ulterioară a acestei teorii. A adus contribuții importante în studiul ecuațiilor cu derivate parțiale, ecuațiilor cu diferențe finite, serii etc. Laplace este autorul unui tratat de mecanică cerească în cinci volume, la care a lucrat mai bine de 25 de ani (1799—1825), operă fundamentală în domeniul astronomiei, în care au fost cuprinse într-un corp de doctrină omogenă toate cercetările întreprinse începînd de la Newton pentru explicarea diferitelor fenomene astronomice cu ajutorul legii gravitației universale.

Scanned with OKEN Scanner

Care sînt trăsăturile caracteristice ale legii normale de probabilitate. O variabilă aleatoare care se supune unei legi normale de probabilitate, ia o infinitate de valori. O lege normală este continuă. Densitatea ei este

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-m)^2}{2\sigma^2}}.$$

După cum se vede, în componența legii normale intră doi parametrii m și σ, a căror semnificație, după cum vom dovedi mai tîrziu, este simplă

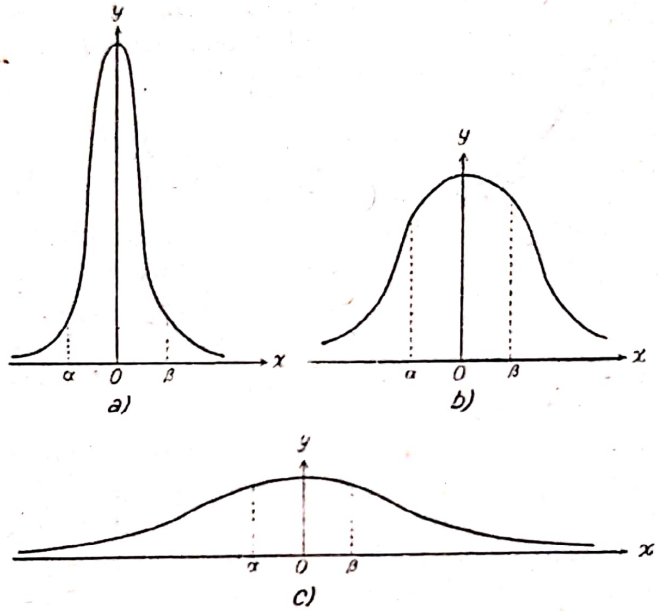


Fig. 7

și anume : m reprezintă valoarea medie a variabilei care admite pe $f(x)$ ca densitate de probabilitate, iar σ dispersia ei. Adică, putem scrie

$$\frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-\frac{(x-m)^2}{2\sigma^2}} dx = 1,$$

$$\frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{+\infty} x e^{-\frac{(x-m)^2}{2\sigma^2}} dx = m,$$

$$\frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{+\infty} (x-m)^2 e^{-\frac{(x-m)^2}{2\sigma^2}} dx = \sigma^2.$$

Legea normală cu densitatea de probabilitate $f(x)$ o vom nota prin $N(m, \sigma)$, pentru a pune în evidență cei doi parametrii de care depinde.

În figura 7 sînt reprezentate cîteva modele teoretice de legi normale de probabilitate, avînd numărul probelor n același.

Pe baza acestor modele se pot observa următoarele proprietăți comune tuturor curbelor normale:

a) Toate curbele sînt simetrice față de axa y-lor care trece prin punctul de maxim. Aceasta înseamnă că toate valorile de pe axa variabilei aleatoare X, egal depărtate de modul — valoarea cea mai probabilă — au probabilități egale.

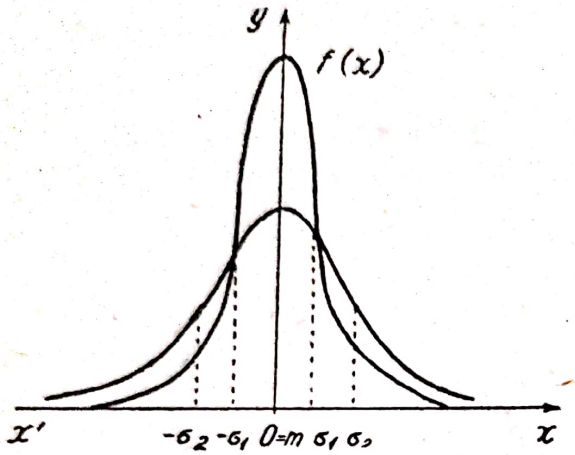


Fig. 8

b) Toate curbele admit un punct de maxim și scad necontenit la dreapta și la stînga lui. Vom vedea că acest punct de maxim reprezintă densitatea de probabilitate a mediei $f(m)$ și că

$$f(m) = \frac{1}{\sigma\sqrt{2\pi}}.$$

c) Toate curbele au formă de clopot, avînd convexitatea îndreptată în sus în apropierea punctului de maxim, iar la o anumită distanță de el, curbele își schimbă semnul convexității; ele au două puncte de inflexiune și aceasta în punctele $X = \pm\sigma$.

În figura 8 proprietățile de mai sus sînt sugestiv reprezentate.

Legea normală, fiind continuă, se înțelege de la sine că nu poate fi o distribuție experimentală, directă, deoarece volumul oricărei distribuții care rezultă din experiențe este finit. S-a constatat însă, că numeroase distribuții statistice experimentale se supun unei legi normale, prin faptul că *frecvențele empirice* sînt foarte apropiate de cele date de această lege *). Printre aceste distribuții se găsește și distribuția binomială.

Am văzut că legea binomială este asimetrică, cu excepția cazului $p = q = 0{,}5$ și că această asimetrie este cu atît mai pronunțată cu cît p este mai mic.

Din exemplul arătat în tabelul 5 se poate observa totuși că dacă n crește, asimetria devine din ce în ce mai puțin pronunțată, deși probabilitatea $p = 0{,}1$.

Constatări de acest fel au sugerat ideea că legea normală de probabilitate poate fi dedusă ca o limită a legii binomiale. Acest lucru a fost făcut pentru prima oară de Moivre **) în 1733.

---

*) Vezi cunoscutul exemplu al lui Quetelet privind distribuția unor recruți în raport cu circumferința toracelui măsurat în țoli. O. Onicescu, Gh. Mihoc, *Lecții de statistică matematică*. Ed. tehnică, București, 1958, p. 154.

**) A. Moivre (1667—1754), a fost profesor de matematică la Londra. A descoperit, în afară de formula legii normale, încă alte două formule fundamentale în matematică și anume expresia asimptotică a lui $n!$ și formula care exprimă ridicarea unui număr complex, pus sub formă trigonometrică, la o putere întreagă.

Tabelul 5

Valorile probabilităţilor 10.000 $P_{n(\alpha)}$ calculate pe baza probabilităţilor p = 0,1 şi q = 0,9, pentru un număr diferit de probe, conform schemei lui J. Bernoulli

α	$n = 5$	$n = 10$	$n = 20$	$n = 50$	$n = 100$
0	5 905	3 487	1 216	51	—
1	3 281	3 874	2 702	286	3
2	729	1 937	2 852	779	16
3	81	574	1 901	1 386	59
4	4	112	898	1 809	159
5	—	14	319	1 849	338
6	—	2	89	1 541	596
7	—	—	20	1 076	889
8	—	—	4	643	1 148
9	—	—	1	333	1 304
10			—	155	1 319
11			—	61	1 199
12			—	22	988
13			—	7	743
14				2	513
15				—	327
16				—	193
17				—	106
18				—	54
19				—	26
20					12
21					5
22					2
23					1

Am arătat că probabilitatea ca din n extracţii succesive dintr-o urnă Bernoulli cu două stări să obţinem α bile albe este dată de legea binomială :

$$P_{n(\alpha)} = C_n^{\alpha} p^{\alpha} q^{n-\alpha} = \frac{n!}{\alpha!\,(n-\alpha)!} \cdot p^{\alpha} q^{n-\alpha}. \qquad \text{(VII,2.1)}$$

Aplicaţiile cele mai importante ale acestei legi sînt date pentru valorile mari ale lui n, adică atunci cînd avem colectivităţi formate dintr-un număr mare de fenomene întîmplătoare identice. Pentru valorile mari ale lui n intervine însă o dificultate de calcul : aceea de a determina $n!$ Această dificultate se poate înlătura folosind formula asimptotică a lui Moivre-Stirling :

$$n! = n^n e^{-n} \sqrt{2\pi n}\,[1 + \theta(n)],$$

în care $\theta(n)$ tinde către zero, cînd n creşte necontenit. Practic, influenţa lui $\theta(n)$ este neglijabilă, aşa că vom folosi formula lui Moivre-Stirling, sub forma

$$n! \cong n^n e^{-n} \sqrt{2\pi n}.$$

Ştim că valoarea medie a variabilei α este egală cu np, dispersia $\sigma^2 = npq$, iar abaterea medie pătratică $\sigma = \sqrt{npq}$.

Să introducem abaterea numărului α de la valoarea lui medie np ;

$$l = \alpha - np \quad \text{şi} \quad \alpha = l + np.$$

Se vede că raportul

$$\beta = \frac{l}{n} = \frac{\alpha - np}{n} = \frac{\alpha}{n} - p$$

ne dă o valoare relativă a abaterii, pe care o vom numi *abatere relativă*.

Abaterile absolute l le putem compara cu abaterea medie pătratică $\sigma = \sqrt{npq}$, în care caz raportul

$$\lambda = \frac{l}{\sigma} = \frac{\alpha - np}{\sqrt{npq}} \qquad \text{(VII,2.2)}$$

ne dă *abaterea relativă redusă* sau pe scurt *abatere redusă*.

Este evident că $P_{n(\alpha)}$, probabilitatea de a obține în n probe de α ori evenimentul E, este tot una cu probabilitatea ca în n probe să obținem abaterea absolută l. Prin urmare, notînd cu $P_{n(l)}$, această probabilitate, avem

$$P_{n(\alpha)} = P_{n(l)}.$$

În formula (VII,2.1) putem deci înlocui pe α prin $l + np$ și pe $P_{n(\alpha)}$ prin $P_{n(l)}$, și se obține

$$P_{n(l)} = \frac{n!}{(np + l)!\,(nq - l)!} \cdot p^{np+l} q^{nq-l}.$$

Folosind acum formula lui Moivre-Stirling și efectuînd simplificările necesare, ajungem la formula

$$P_{n(l)} = \frac{1}{\sqrt{2\pi npq}\left(1 + \frac{l}{np}\right)^{\left(np+l+\frac{1}{2}\right)}\left(1 - \frac{l}{nq}\right)^{nq-l+\frac{1}{2}}} \qquad \text{(VII,2.3)}$$

Notăm

$$M = \left(1 + \frac{l}{np}\right)^{np+l+\frac{1}{2}}\left(1 - \frac{l}{nq}\right)^{nq-l+\frac{1}{2}}.$$

Pentru evaluarea asimptotică a lui M, vom aplica logaritmii și vom avea

$$\lg M = \left(np + l + \frac{1}{2}\right)\lg\left(1 + \frac{l}{np}\right) + \left(nq - l + \frac{1}{2}\right)\lg\left(1 - \frac{l}{nq}\right).$$

Ne propunem să evaluăm asimptotic pe $P_{n(l)}$, pentru valorile lui l îndeplinind condiția $|l| < A\sqrt{n}$, unde A este un număr pozitiv dat.

În acest caz

$$\frac{|l|}{np} < 1, \quad \frac{|l|}{nq} < 1,$$

pentru valorile lui n suficient de mari. Avem, prin urmare, dezvoltările în serie

$$\lg\left(1 + \frac{l}{np}\right) = \frac{l}{np} - \frac{l^2}{2n^2p^2} + \frac{l^3}{3n^3p^3} - \dots$$

$$\lg\left(1 - \frac{l}{nq}\right) = -\frac{l}{nq} - \frac{l^2}{2n^2q^2} - \frac{l^3}{3n^3q^3} - \dots$$

Scanned with OKEN Scanner

Introducîndu-le în relaţia care dă pe $\lg M$ şi ţinînd seama de faptul că $|l| < A\sqrt{n}$, obţinem următoarea evaluare asimptotică:

$$\lg M = -\frac{l^2}{2npq} + \theta\left(\frac{1}{\sqrt{n}}\right),$$

în care notăm cu $\theta\left(\frac{1}{\sqrt{n}}\right)$ o funcţie de n care tinde către zero cînd n tinde către infinit.

Pentru valorile destul de mari ale lui n, putem deci scrie

$$M \approx e^{+\frac{l^2}{2npq}}$$

şi prin urmare, ţinînd seama de (VII,2.3), avem

$$P_{n(l)} \simeq \frac{1}{\sqrt{2\pi npq}} e^{-\frac{l^2}{2npq}}. \qquad \text{(VII,2.4)}$$

Este util de observat că dacă notăm

$$\sigma = \sqrt{npq}$$

formula (VII,2.4) devine

$$P_{n(l)} = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{l^2}{2\sigma^2}} \qquad \text{(VII,2.5)}$$

care este tocmai *legea normală de probabilitate*.

Formulele (VII,2.4) şi (VII,2.5) ne arată că probabilitatea ca din n probe să obţinem o abatere l, în condiţiile date, poate fi pusă asimptotic sub forma unei legi normale.

Ne interesează însă mai mult decît atît şi anume să evaluăm asimptotic probabilitatea de a obţine o abatere l cuprinsă într-un anumit interval şi anume:

$$-\lambda'\sqrt{2npq} < l < \lambda''\sqrt{2npq}, \qquad \text{(VII,2.6)}$$

unde λ' şi λ'' sînt două numere finite date.

Considerăm din nou probabilitatea $P_{n(\alpha)}$ ca din n extracţii să obţinem α bile albe:

$$P_{n(\alpha)} = \frac{n!}{\alpha!\,(n-\alpha)!} p^{\alpha} q^{n-\alpha}.$$

Notăm

$$\alpha = np + x\sqrt{2npq}, \qquad \text{(VII,2.7)}$$

de unde rezultă

$$x = \frac{\alpha - np}{\sqrt{2npq}} = \frac{l}{\sqrt{2npq}}; \qquad \text{(VII,2.8)}$$

x poartă numele de abatere relativă.

Vom repeta acelaşi calcul de aproximare de mai sus, aplicînd formula lui Moivre-Stirling probabilităţii $P_{n(\alpha)}$, după ce am substituit pe α prin valoarea dată de (VII,2.7). În baza acestui calcul, care este inutil să-l mai redăm aici, deoarece este de aceeaşi natură cu cel care ne-a condus la formula (VII,2.4), obţinem

$$P_{n(k)} = \frac{e^{-x^2}}{\sqrt{2\pi npq}} [1 + \xi(n, \alpha)], \qquad \text{(VII,2.9)}$$

unde pentru valorile lui x și α, îndeplinind condiția (VII, 2.6), adică

$$-\lambda' < x < \lambda''$$

$$np - \lambda' \sqrt{2npq} < \alpha < np + \lambda'' \sqrt{2npq},$$

$\xi(n, \alpha)$ este o funcție de n și α, care tinde uniform către zero, cînd n și α tind către infinit.

Să notăm prin $P_{(n\lambda',\lambda'')}$ probabilitatea inegalităților

$$-\lambda' \sqrt{2npq} < \alpha - np < \lambda'' \sqrt{2npq}$$

și prin α' și α'' respectiv cea mai mică și cea mai mare valoare a lui α, care verifică aceste inegalități.

Avem

$$P_{n(\lambda',\lambda'')} = \sum_{\alpha=\alpha'}^{\alpha''} P_{n(\alpha)}$$

sau, ținînd seama de (VII,2.9),

$$P_{n(\lambda',\lambda'')} = \sum_{\alpha=\alpha'}^{\alpha''} \frac{e^{-x^2}}{\sqrt{2\pi npq}} [1 + \xi(n, \alpha)] = \sum_{\alpha=\alpha'}^{\alpha''} \frac{e^{-x^2}}{\sqrt{2\pi npq}} + \sum_{\alpha'}^{\alpha''} \frac{e^{-x^2}}{\sqrt{2\pi npq}} \xi(n, \alpha). \qquad \text{(VII,2.10)}$$

Să mai observăm că din (VII,2.8) rezultă că x este o funcție de α și că avem

$$x(\alpha + 1) - x(\alpha) = \frac{1}{\sqrt{2npq}} > 0. \qquad \text{(VII,2.11)}$$

Egalitatea (VII,2.10) poate fi deci scrisă sub forma

$$P_{n(\lambda',\lambda'')} = \frac{1}{\sqrt{\pi}} \sum_{\alpha=\alpha'}^{\alpha''} e^{-x^2(\alpha)} [x(\alpha + 1) - x(\alpha)] + T_{n(\alpha)}, \qquad \text{(VII,2.12)}$$

unde prin $T_{n(\alpha)}$ am însemnat termenul

$$T_{n(\alpha)} = \sum_{\alpha=\alpha'}^{\alpha''} \frac{e^{-x^2(\alpha)}}{\sqrt{2\pi npq}} \xi(n, \alpha).$$

Avem însă

$$|T_{n(\alpha)}| = \left| \sum_{\alpha=\alpha'}^{\alpha''} \frac{e^{-x^2(\alpha)}}{\sqrt{2\pi npq}} \xi(n, \alpha) \right| \leqslant \sum_{\alpha=\alpha'}^{\alpha''} |\xi(n, \alpha)| \frac{e^{-x^2(\alpha)}}{\sqrt{2\pi npq}} < \theta |\bar{\xi}(n, \alpha)| \sum_{\alpha=\alpha'}^{\alpha''} \frac{e^{-x^2(\alpha)}}{\sqrt{2\pi npq}},$$

unde θ este un număr cuprins între 0 și 1, iar $|\bar{\xi}(n, \alpha)|$ cea mai mare dintre valorile $|\xi(n, \alpha)|$, în intervalul $[\alpha', \alpha'']$.

Ținînd seama de (VII,2.11), din inegalitățile de mai sus deducem mai departe

$$|T_{n(\alpha)}| < \frac{\theta}{\sqrt{\pi}} |\bar{\xi}(n, \alpha)| \sum_{\alpha=\alpha'}^{\alpha''} [x(\alpha + 1) - x(\alpha)] e^{-x^2(\alpha)}. \qquad \text{(VII,2.13)}$$

Să considerăm suma

$$\sum_{\alpha=\alpha'}^{\alpha''} |x(\alpha + 1) - x(\alpha)| e^{-x^2(\alpha)}. \qquad \text{(VII,2.14)}$$

În baza egalităților (VII,2.11), punctele $x(\alpha)$ împart segmentul $(-\lambda', \lambda'')$ în intervale egale, care tind către zero, cînd n crește necontenit. Expresia (VII,2.14) tinde deci, pentru

$n \to \infty$, către

$$\frac{1}{\sqrt{\pi}} \int_{-\lambda'}^{\lambda''} e^{-x^2} dx.$$

Pe de altă parte, deoarece

$$\lim_{n\to\infty} |\bar{\xi}(n, \alpha)| = 0,$$

rezultă din (VII,2.13) că

$$\lim_{n\to\infty} T_{n(\alpha)} = 0.$$

Ţinînd seama de aceste rezultate, din (VII,2,12) se obţine

$$\lim_{n\to\infty} P_{n(\lambda',\lambda'')} = \frac{1}{\sqrt{\pi}} \int_{-\lambda'}^{+\lambda''} e^{-x^2} dx$$

sau

$$\lim_{n\to\infty} P[np - \lambda' \sqrt{2npq} < \alpha < np + \lambda'' \sqrt{2npq}] = \frac{1}{\sqrt{\pi}} \int_{-\lambda'}^{+\lambda''} e^{-x^2} dx \qquad \text{(VII,2.15)}$$

Proprietatea de limită pe care am demonstrat-o, mai poate fi scrisă sub formă echivalentă

$$\lim_{n\to\infty} P\left[-\lambda' < \frac{\alpha - np}{\sqrt{2npq}} < \lambda''\right] = \frac{1}{\sqrt{\pi}} \int_{-\lambda'}^{+\lambda''} e^{-x^2} dx. \qquad \text{(VII,2.15')}$$

Din relaţia (VII, 2.15') reiese că funcţia de repartiţie a variabilei

$$x = \frac{\alpha - np}{\sqrt{2npq}}$$

tinde, pentru $n \to \infty$, către funcţia de repartiţie a variabilei normale cu densitatea

$$\frac{1}{\sqrt{\pi}} e^{-x^2}. \qquad \text{(VII,2.16)}$$

Comparînd această lege normală cu legea normală $N(m, \sigma)$ care are densitatea

$$\frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-m)^2}{2\sigma^2}},$$

se vede că pentru legea normală (VII,2.16) avem $m = 0$ şi $\sigma = \frac{1}{\sqrt{2}}$. Prin urmare, densitatea (VII,2.16) corespunde unei legi normale $N\left(0, \frac{1}{\sqrt{2}}\right)$.

Dacă în relaţia (VII,2.15') punem $\lambda' = \frac{\mu'}{\sqrt{2}}$, $\lambda'' = \frac{\mu''}{\sqrt{2}}$, obţinem,

$$\lim_{n\to\infty} P\left[-\mu' < \frac{\alpha - np}{\sqrt{npq}} < \mu''\right] = \frac{1}{\sqrt{\pi}} \int_{-\frac{\mu'}{\sqrt{2}}}^{\frac{\mu''}{\sqrt{2}}} e^{-x^2} dx.$$

Efectuînd în integrala din membrul al doilea substituția $x = \frac{y}{\sqrt{2}}$, se obține

$$\frac{1}{\sqrt{2\pi}} \int_{-\mu'}^{\mu''} e^{-\frac{x^2}{2}} dx.$$

Putem deci scrie

$$\lim_{n\to\infty} P\left\{-\mu' < \frac{\alpha - np}{\sqrt{npq}} < \mu''\right\} = \frac{1}{\sqrt{2\pi}} \int_{-\mu'}^{\mu''} e^{-\frac{x^2}{2}} dx. \qquad \text{(VII,2.17)}$$

Relația (VII,2.17), ne arată că funcția de repartiție a abaterii reduse

$$x = \frac{\alpha - np}{\sqrt{npq}}$$

tinde către funcția de repartiție a legii normale $N(0,1)$.

Atît relația (VII,2.15′) cît și relația (VII,2.17) sînt importante prin faptul că legile limită din membrul al doilea ale acestor egalități, anume $N\left(0, \frac{1}{\sqrt{2}}\right)$ și $N(0,1)$, nu depind de p.

Dacă se calculează o tabelă de valori pentru una din legile normale $N\left(0, \frac{1}{\sqrt{2}}\right)$ și $N(0,1)$, această tabelă poate fi utilizată pentru determinarea asimptotică a probabilității de a avea o abatere redusă într-un interval dat pentru *toate distribuțiile binomiale.*

Cu privire la relațiile (VII,2.15′) și (VII,2.17) mai trebuie să facem o observație esențială. În ambele relații avem în membrul întîi un proces de limită. În aplicațiile, de orice natură, în care întîlnim distribuția binomială, numărul experiențelor este întotdeauna finit. Nu există deci o aplicație practică în care să se realizeze procesul de limită din egalitățile sus-menționate. În baza acestui proces putem însă scrie

$$P\left\{-\mu' < \frac{\alpha - np}{\sqrt{npq}} < \mu''\right\} \simeq \frac{1}{\sqrt{2\pi}} \int_{-\mu'}^{\mu''} e^{-\frac{x^2}{2}} dx, \qquad \text{(VII,2.18)}$$

adică pentru valori mari ale lui n, probabilitatea din membrul întîi, care ne este necunoscută, este foarte apropiată ca valoare de probabilitatea cunoscută din membrul al doilea, luată din tabele.

Întotdeauna cînd aplicăm relația (VII,2.18), trebuie să ținem seama de faptul că această relație are numai un caracter asimptotic, constituind o aproximație din ce în ce mai bună pentru probabilitatea necunoscută din membrul întîi, cu cît numărul probelor n este mai mare.

Să notăm

$$H(\lambda) = \frac{1}{\sqrt{2\pi}} \int_0^{\lambda} e^{-\frac{x^2}{2}} dx.$$

Valorile lui $H(\lambda)$, pentru diferitele valori pozitive date lui λ, se găsesc în tabelul anexă I. Datorită simetriei, avem pentru $\lambda > 0$:

$$H(\lambda) = \frac{1}{\sqrt{2\pi}} \int_{\lambda}^{0} e^{-\frac{x^2}{2}} dx.$$

Deducem de aici că dacă $\mu' > 0$, $\mu'' > 0$, putem scrie

$$\frac{1}{\sqrt{2\pi}} \int_{-\mu'}^{\mu''} e^{-\frac{x^2}{2}} dx = H(\mu') + H(\mu'').$$

Dacă $\mu' < 0$, $\mu'' > 0$, $\mu'' > -\mu'$ rezultă

$$\frac{1}{\sqrt{2\pi}} \int_{-\mu'}^{\mu''} e^{-\frac{x^2}{2}} dx = H(\mu'') - H(-\mu').$$

Cu alţi termeni, pentru folosirea tabelei $H(\lambda)$, (tabela anexă I), egalitatea asimptotică (VII, 2.18) poate fi scrisă, ţinînd seama de consideraţiile de mai sus, sub una din formele următoare:

$$P\left[-\lambda' < \frac{\alpha - np}{\sqrt{npq}} < \lambda''\right] \simeq H(\lambda') + H(\lambda''), \quad [\lambda' > 0, \ \lambda'' > 0],$$

$$P\left[\lambda' < \frac{\alpha - np}{\sqrt{npq}} < \lambda''\right] \simeq H(\lambda'') - H(\lambda'), \quad [\lambda'' > \lambda' > 0], \qquad \text{(VII, 2.19)}$$

$$P\left[-\lambda' < \frac{\alpha - np}{\sqrt{npq}} < -\lambda''\right] \simeq H(\lambda') - H(\lambda''), \quad [\lambda' > \lambda'' > 0].$$

Folosirea în practică a formulelor (VII,2.19) nu este indicată înainte de a cunoaşte gradul lor de aproximare. Dacă procesul de convergenţă din membrul întîi al egalităţior (VII,2.19) este foarte încet, putem comite erori mari folosindu-le pentru anumite valori ale lui n. Pentru înlăturarea acestei deficienţe s-au publicat multe lucrări.

Vom c ta în această direcţie următoarea teoremă, datorită lui S. Bernstein *), pe care o enunţăm folosind notaţiile pînă acum întrebuinţate:

TEOREMĂ. *Pentru*

$$n \geqslant \frac{62{,}5}{pq} \quad \text{şi} \quad 0 < \lambda < \sqrt[6]{16npq}$$

avem

$$P\left[-\lambda + \frac{q-p}{6\sqrt{npq}}\lambda^2 < \frac{\alpha - np}{\sqrt{npq}} < \lambda + \frac{q-p}{6\sqrt{npq}}\lambda^2\right] = 2H(\lambda) + R_n,$$

unde

$$|R_n| < 2\, e^{-\sqrt{2npq}}.$$

Teorema lui Bernstein ne dă o limitare precisă asupra erorii pe care o facem, cînd folosim anumite valori ale lui n şi λ, îndeplinind condiţiile cerute. Se observă că condiţia impusă lui λ este aproape întotdeauna îndeplinită pentru valorile mari ale lui n, deoarece nu ne interesează valorile lui λ depăşind ordinul unităţilor.

Ca să ne dăm seama de modul de aplicare al teoremei lui Bernstein vom considera un exemplu, alegînd $p = 0{,}6$ şi $q = 0{,}4$.

*) Serghei Natanovici Bernstein (n. 1880), matematician sovietic, cunoscut prin lucrările sale de analiză matematică şi calculul probabilităţilor. A publicat lucrări importante asupra aproximării funcţiilor şi a extins teorema lui Liapunov pentru categorii foarte generale de variabile dependente.

Scanned with OKEN Scanner

În acest caz prima condiție ne dă

$$n \geqslant \frac{62{,}5}{pq}, \qquad n \geqslant 261.$$

Formula este deci aplicabilă pentru un număr mai mare de 260 de experiențe.
Pentru λ avem delimitarea

$$0 < \lambda < \sqrt[6]{16npq}, \qquad 0 < \lambda < \sqrt[6]{\frac{96n}{25}}.$$

Luînd un număr mic de experiențe, $n = 300$, rezultă mai departe:

$$0 < \lambda < 2\sqrt[6]{18}, \qquad 0 < \lambda < 3.25.$$

Putem deci lua $\lambda = 2$.
Din inegalitățile

$$-\lambda + \frac{q-p}{6\sqrt{npq}}\lambda^2 < \frac{\alpha - np}{\sqrt{npq}} < \lambda + \frac{q-p}{6\sqrt{npq}}\lambda^2$$

deducem pentru $p = 0{,}6$, $n = 300$, $\lambda = 2$,

$$163 < \alpha < 196{\cdot}8.$$

Însă,

$$P[163 < \alpha < 196{,}8] = P[163 < \alpha < 197],$$

deoarece α nu poate lua decît valori întregi.

Pe de altă parte,

$$2H(2) = 0{,}9555,$$

iar

$$|R_{300}| < 2\,e^{-12} < \frac{1}{150\,000}.$$

Prin urmare, făcînd 300 de experiențe, numărul α de apariții a bilei albe, dacă $p = 0{,}6$, este cuprins între 167 și 192 cu o probabilitate egală cu 0,955. Eroarea pe care o comitem folosind această probabilitate este mai mică decît 0,00001, neafectînd deci ultima zecimală din probabilitatea 0,9555.

Teorema lui Bernstein este aplicabilă pentru valori mici ale lui n, cînd p și q sînt foarte apropiate între ele. Dacă însă luăm $p = 0{,}01$, $q = 0{,}99$, din formula

$$n \geqslant \frac{62{,}5}{pq}$$

deducem

$$n \geqslant 63\,131.$$

Deci numai după un număr foarte mare de experiențe formula de aproximație dată poate fi aplicată. Cu cît probabilitatea p este mai mică, cu atît convergența probabilității $P_{n(\alpha)}$ către legea normală este mai lentă, după cum vom vedea atunci cînd vom examina distribuția Poisson.

Pentru folosirea formulelor (VII, 2.19) este util să ținem seama de mărimea erorilor pe care le comitem pentru cîteva valori mai semnificate date lui λ' și λ''.

Astfel, dacă în prima din aceste formule punem

$$\lambda' = \lambda'' = k\sqrt{2},$$

pentru valorile lui k:

$$\frac{\sqrt{npq}}{4}, \quad \frac{\sqrt{npq}}{2}, \quad \sqrt{npq}, \quad \frac{3\sqrt{npq}}{2}, \quad 2\sqrt{npq}, \quad \frac{5\sqrt{npq}}{2}, \quad 3\sqrt{npq},$$

corespund respectiv erorile aproximative:

$$\frac{1,96}{\sqrt{npq}}, \quad \frac{0,92}{\sqrt{npq}}, \quad \frac{0,35}{\sqrt{npq}}, \quad \frac{0,15}{\sqrt{npq}}, \quad \frac{0,056}{\sqrt{npq}}, \quad \frac{0,018}{\sqrt{npq}}, \quad \frac{0,004}{\sqrt{npq}}. \qquad \text{(VII,2.20)}$$

Sîntem acum în măsură, folosind legea normală, în baza formulelor (VII,2.19), să rezolvăm unele probleme, importante din punct de vedere statistic, relative la distribuția binomială. Vom împărți aceste probleme în mai multe categorii:

I. Determinarea probabilității obținerii unei abateri absolute cuprinse între $-h$ și h, după n probe consecutive

În acest caz, trebuie să avem inegalitățile

$$-h < \alpha - np < h,$$

care se pot pune sub forma

$$-\frac{h}{\sqrt{npq}} < \frac{\alpha - np}{\sqrt{npq}} < \frac{h}{\sqrt{npq}}.$$

În baza primei dintre egalitățile asimptotice (VII,2.19), înseamnă că

$$\lambda' = \lambda'' = \frac{h}{\sqrt{npq}}$$

și că probabilitatea căutată este aproximativ egală cu $2H\left(\frac{h}{\sqrt{npq}}\right)$.

Pentru determinarea erorii pe care o comitem folosind această probabilitate, ne orientăm după valorile (VII,2.20).

EXEMPLUL 29. Să se determine probabilitatea ca aruncînd o monedă de 1600 ori, o față anumită să apară de un număr de ori cuprins între 755 și 845.

Avem

$$p = q = \frac{1}{2}, \quad n = 1\,600, \quad 755 < \alpha < 845.$$

Cum $np = 800$, înseamnă că

$$-45 < \alpha - np < 45,$$

Avînd $\sqrt{npq} = 20$, prin împărțirea inegalităților precedente cu 20, obținem

$$-2{,}25 < \frac{\alpha - np}{\sqrt{npq}} < 2{,}25.$$

Folosim prima dintre egalitățile de mai sus, luînd $\lambda' = \lambda'' = 2{,}25$. Probabilitatea căutată este

$$2H(2{,}25) = 0{,}97556.$$

Din (VII,2.20) deducem că pentru $k = 30$ corespunde o eroare egală cu 0,075 și pentru $k = 40$ o eroare egală cu 0,0028. În cazul nostru avem

$$k = \frac{\lambda}{\sqrt{2}} = \frac{45}{\sqrt{2}} = 31{,}82.$$

Procedînd prin interpolare găsim că prin folosirea probabilității $2H(2{,}25)$, comitem o eroare aproximativ egală cu 0,0024. Se vede că nu putem conta decît pe primele două zecimale. Probabilitatea căutată este egală cu 0,97, ambele zecimale fiind exacte.

II. Determinarea probabilității obținerii unei abateri relative cuprinse între — θ și θ, după n probe consecutive

EXEMPLUL 30. Dintr-o urnă care conține același număr de bile albe și negre se fac 2 500 extracții succesive, punîndu-se după fiecare extracție bila înapoi în urnă, și se împarte numărul de bile albe obținute la 2 500. Care este probabilitatea ca frecvența astfel obținută să difere de 1/2 cu mai puțin de 0,02?

În acest caz, avem

$$n = 2\,500, \quad p = \frac{1}{2}, \quad \sqrt{npq} = 25.$$

Trebuie să găsim probabilitatea inegalităților

$$-0{,}02 < \frac{\alpha}{n} - p < 0{,}02,$$

care pot fi scrise sub forma

$$-2 < \frac{\alpha - np}{\sqrt{npq}} < 2.$$

În baza formulelor (VII,2.19) această probabilitate este egală cu

$$2H(2) = 0{,}955.$$

EXEMPLUL 31. În aceleași condiții ca în exemplul 30, să se găsească probabilitatea de a obține o abatere relativă cuprinsă între $-0,03$ și $-0,01$.

Trebuie să avem

$$-0,03 < \frac{\alpha}{n} - p < -0,01$$

sau

$$-3 < \frac{\alpha - np}{\sqrt{npq}} < -1.$$

În baza ultimei dintre relațiile (VII, 2.19), această probabilitate este

$$H(3) - H(1) = 0,29865 - 0,14134 = 0,15731.$$

III. Determinarea numărului de probe necesare pentru a obține cu o probabilitate dată o abatere cuprinsă într-un interval dat

EXEMPLUL 32. De cîte ori trebuie să aruncăm un zar pentru ca probabilitatea ca frecvența relativă să fie cuprinsă între $-0,03$ și $0,03$ să fie egală cu $0,95$?

În cazul nostru trebuie ca probabilitatea inegalităților

$$-0,03 < \frac{\alpha}{n} - p < 0,03$$

să fie egală cu $0,95$, știind că $p = \frac{1}{6}$.

Inegalitățile de mai sus pot fi scrise sub forma

$$-\frac{0,03n}{\sqrt{npq}} < \frac{\alpha - np}{\sqrt{npq}} < \frac{0,03n}{\sqrt{npq}}.$$

În baza primei dintre egalitățile (VII,2.19) trebuie să avem

$$2H\left(\frac{0,03n}{\sqrt{npq}}\right) = 0,95.$$

Din tabelul anexă I deducem

$$\frac{0,03n}{\sqrt{npq}} = 1,96.$$

Deoarece $p = \frac{1}{6}$, rezultă

$$n = 1\,334.$$

Dacă vom face mai mult de 1 334 probe, probabilitatea ca să avem o abatere relativă cuprinsă între $-0,03$ și $0,03$ este mai mare de $0,95$.

Scanned with OKEN Scanner

TEOREMA CENTRALĂ A CALCULULUI PROBABILITĂȚILOR. Am văzut că funcția de repartiție a abaterii reduse

$$x = \frac{\alpha - np}{\sqrt{npq}}$$

tinde, pentru $n \to \infty$, către funcția de repartiție a legii normale.

Dacă notăm prin ξ_i numărul de bile albe obținut în extracția de rang i, număr care nu poate fi decît 1 sau 0, numărul α de bile obținut în n probe este

$$\alpha = \xi_1 + \xi_2 + \ldots + \xi_n .$$

În cazul nostru variabilele ξ_i sînt :

$$\xi_i : \begin{pmatrix} 1 & 0 \\ p & q \end{pmatrix}$$

adică niște variabile care cu probabilitatea p și q iau respectiv valorile 1 și 0. Sînt deci niște variabile particulare.

Se pune problema dacă proprietatea semnalată (anume suma variabilelor ξ_i, convenabil normată, tinde către legea normală) nu se păstrează și pentru variabile ξ_i mai generale decît cele de mai sus.

Cel dintîi care s-a ocupat de această problemă a fost, după cum am mai arătat, Laplace. Pe baza a numeroase experiențe, s-a constatat că erorile de observație se distribuie aproximativ după o lege normală. Era necesar să se dea o explicație teoretică acestei constatări experimentale. Laplace a avut intuiția că eroarea finală a unor operații de măsurare este suma unui mare număr de erori parțiale și că legea normală este consecința acestui fapt.

Teorema care arată condițiile în care suma unui număr de variabile aleatoare independente converge către legea normală poartă numele de *teorema centrală din calculul probabilităților*.

Fie

$$\xi_{(n)} = \xi_1 + \xi_2 + \ldots + \xi_n .$$

Se poate arăta că în condiții destul de generale variabila $\xi_{(n)}$, convenabil normată, tinde către legea normală. Aceste condiții au fost date de Liapunov *). Condițiile stabilite de Liapunov sînt atît de generale, încît justifică faptul că legea normală este atît de mult întîlnită în aplicațiile statistice. Vom nota :

$$m_k = M(\xi_k),$$

$$\sigma_k^2 = M(\xi_k - m_k)^2,$$

$$\rho_k^3 = M(|\xi_k - m_k|^3),$$

unde $k = 1,2,\ldots,n$.

*) A. M. Liapunov (1857—1918), este cunoscut prin contribuția importantă pe care a adus-o prin lucrările lui în teoria probabilităților și mecanică. El a elaborat metoda generală pentru rezolvarea problemelor de stabilitate în mecanică. Generalizînd rezultatele obținute de Cebîșev și Markov, Liapunov a demonstrat teorema centrală din teoria probabilităților în condiții cu mult mai generale decît predecesorii săi. Remarcabil e faptul că a folosit pentru prima oară metoda funcției caracteristice.

Scanned with OKEN Scanner

Să însemnăm mai departe:

$$\sigma_{(n)}^2 = \sigma_1^2 + \sigma_2^2 + \ldots + \sigma_n^2,$$

$$\rho_{(n)}^3 = \rho_1^3 + \rho_2^3 + \ldots + \rho_n^3.$$

Teorema lui Liapunov afirmă că dacă

$$\lim_{n\to\infty} \frac{\rho_{(n)}}{\sigma_{(n)}} = 0$$

variabila $\xi_{(n)}$, normată, tinde la limită, către legea normală.

Dacă notăm:

$$m = M(\xi),$$
$$\sigma^2 = M(\xi - m)^2$$

avem, în baza proprietăților de adunare ale valorilor medii,

$$m = \sum_{k=1}^{n} m_k,$$

$$\sigma_{(n)}^2 = \sum_{k=1}^{n} \sigma_k^2.$$

Fie $\varphi_k(t)$ funcția caracteristică a variabilei $(\xi_k - m_k)$ și $\varphi_{(n)}(t)$ funcția caracteristică a variabilei $\frac{\xi - m}{\sigma}$.

Să considerăm variabila

$$\eta = \frac{\xi - m}{\sigma_{(n)}} = \frac{\sum_{k=1}^{n} (\xi_k - m_k)}{\sigma_{(n)}}.$$

Funcția caracteristică a variabilei $\frac{\xi_k - m_k}{\sigma}$ este $\varphi_k\left(\frac{t}{\sigma}\right)$.

Ținînd seama de relația precedentă, în baza faptului că funcția caracteristică a unei sume finite de variabile independente este egală cu produsul funcțiilor caracteristice ale variabilelor componente ale sumei, rezultă egalitatea

$$\varphi_{(n)}(t) = \varphi_1\left(\frac{t}{\sigma_{(n)}}\right) \varphi_2\left(\frac{t}{\sigma_{(n)}}\right) \ldots \varphi_n\left(\frac{t}{\sigma_{(n)}}\right).$$

Putem scrie însă în general

$$\varphi_k\left(\frac{t}{\sigma_{(n)}}\right) = 1 - \frac{1}{2} \frac{\sigma_k^2}{\sigma_{(n)}^2} t^2 + \frac{1}{6} \theta \frac{\rho_k^3}{\sigma_{(n)}^3} t^3,$$

unde θ este un număr în valoare absolută mai mic decît 1. Mai departe, notăm

$$\lg \varphi_k\left(\frac{t}{\sigma_{(n)}}\right) = \lg(1 + y),$$

unde

$$y = -\frac{\sigma_k^2 t^2}{2\sigma_{(n)}^2} + \theta \frac{\rho_k^3 t^3}{6\sigma_{(n)}^3}.$$

Pe de o parte, din proprietățile generale ale momentelor *) avem $\sigma_k \leqslant \rho_k$, iar pe de altă parte

$$\frac{\rho_k}{\sigma_{(n)}} < \frac{\rho_{(n)}}{\sigma_{(n)}} < 1.$$

Aceasta înseamnă că putem scrie pe y sub forma

$$y = \theta \frac{\rho_k^2}{\sigma_{(n)}^2}\left[\frac{t^2}{2} + \frac{|t|^3}{6}\right], \qquad |\theta| < 1.$$

Pentru $n \to \infty$, conform condiției lui Liapunov, $\frac{\rho_k}{\sigma_{(n)}}$ tinde către zero. Putem alege atunci pe n destul de mare, ca să avem $|y| < \frac{1}{2}$.

Însă

$$\lg(1+y) = y - \frac{y^2}{2} + \frac{y^2}{3} - \frac{y^2}{4} + \ldots = y - \frac{y^2}{2}\left[1 - \frac{2}{3}y + \frac{2}{4}y^2 - \ldots\right] =$$

$$= y + \frac{1}{2}\theta \cdot y^2\left(1 + \frac{1}{2} + \frac{1}{2^2} + \right) = y + \theta y^2.$$

Prin urmare

$$\lg \varphi_k\left(\frac{t}{\sigma_{(n)}}\right) = -\frac{\sigma_k^2}{\sigma_{(n)}^2}\frac{t^2}{2} + \theta \frac{\rho_k^3}{\sigma_{(n)}^3}\left[\frac{1}{6}|t|^3 + \left(\frac{1}{2}t^2 + \frac{1}{6}|t|^3\right)^2\right].$$

Însă

$$\lg \varphi_{(n)}(t) = \sum_{k=1}^{n} \lg \varphi_k\left(\frac{t}{\sigma_{(n)}}\right)$$

și deci, ținînd seama de valoarea lui $\lg \varphi_k\left(\frac{t}{\sigma_{(n)}}\right)$,

$$\lg \varphi_{(n)}(t) = -\frac{t^2}{2} + \theta \frac{\rho_{(n)}^3}{\sigma_{(n)}^3}\left[\frac{1}{6}|t|^3 + \left(\frac{1}{2}t^2 + \frac{1}{6}|t|^3\right)^2\right].$$

Cînd $n \to \infty$, în baza condiției lui Liapunov rezultă

$$\lim_{n\to\infty} \lg \varphi_{(n)}(t) = -\frac{t^2}{2}.$$

Putem scrie

$$\lim_{t\to\infty} \varphi_n(t) = e^{-\frac{t^2}{2}}.$$

Aceasta este însă tocmai funcția caracteristică a legii normale $N(0,1)$.

Funcția caracteristică a variabilei n tinde deci către funcția caracteristică a legii normale, ceea ce demonstrează teorema.

*) Vezi capitolul IV.

Teorema lui Liapunov cuprinde un mare grad de generalitate. Schema lui Bernoulli este inclusă ca un caz particular acestei teoreme.

Într-adevăr, pentru urna lui Bernoulli, avem

$$\sigma_k^2 = pq,$$

$$\rho_k^3 = pq\,(1 - 2p),$$

$$\sigma_{(n)}^2 = npq,$$

$$\rho_{(n)}^3 = npq(1 - 2pq),$$

$$\frac{\rho_{(n)}}{\sigma_{(n)}} = \frac{1}{\sqrt[6]{n}} \cdot \frac{\sqrt[3]{1-2pq}}{\sqrt[6]{pq}}.$$

Prin urmare

$$\lim_{n\to\infty} \frac{\rho_{(n)}}{\sigma_{(n)}} = 0$$

şi deci condiţia lui Liapunov este îndeplinită. Rezultă astfel în mod riguros că abaterile reduse din schema lui Bernoulli se distribuie la limită, cînd numărul probelor creşte necontenit, după legea normală.

Mai mult decît atît, dacă

$$\xi_1, \xi_2, \ldots, \xi_n, \ldots$$

reprezintă un şir de variabile independente avînd aceeaşi lege de probabilitate şi dacă presupunem că există momentele centrate de primele două ordine şi momentul centrat absolut de ordinul al treilea, condiţia lui Liapunov este îndeplinită şi deci suma

$$\xi_1 + \xi_2 + \ldots + \xi_n$$

este asimptotic normală.

Într-adevăr, în acest caz

$$\sigma_{(n)}^2 = n\,\sigma^2, \qquad (\sigma \neq 0),$$

$$\rho_{(n)}^3 = n\rho^3.$$

Prin urmare

$$\lim_{n\to\infty} \cdot \frac{\rho_{(n)}}{\sigma_{(n)}} = \lim \frac{\rho}{\sqrt[6]{n}\cdot\sigma} = 0.$$

Proprietăţi ale legii normale

A. *Momentele centrate ale legii normale* $N(m, \sigma)$ *sînt*:

$$m_{2r} = 1\cdot 3\cdot 5 \ldots (2r - 1)\,\sigma^{2r} = \frac{(2r)!}{2^r \cdot r!}\sigma^{2r}$$

$$m_{2r+1} = 0.$$

Avem, evident,

$$m_{2r+1} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} x^{2r+1} e^{-\frac{x^2}{2}}\,dx = 0$$

deoarece funcția $x^{2r+1} \cdot e^{-\frac{x^2}{2}}$ este o funcție impară și

$$m_{2r} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} x^{2r} e^{-\frac{x^2}{2}} dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} x^{2r-1} d\left(-e^{-\frac{x^2}{2}}\right) =$$

$$= \frac{1}{\sqrt{2\pi}} \left[\left(-x^{2r-1} e^{-\frac{x^2}{2}} \right)_{-\infty}^{+\infty} + (2r-1) \int_{-\infty}^{+\infty} x^{2r-2} e^{-\frac{x^2}{2}} dx \right] = (2r-1) m_{2r-2}.$$

Dînd lui r valorile

$$r,\ r-1,\ \ldots,\ 2,1$$

rezultă următoarele relații de recurență:

$$m_{2r} = (2r-1)\, m_{2r-2},$$

$$m_{2r-2} = (2r-3)\, m_{2r-4},$$

$$m_{2r-4} = (2r-5)\, m_{2r-6},$$

$$\ldots\ldots\ldots\ldots$$

$$m_4 = 3m_2,$$

$$m_2 = m_0, \qquad (m_0 = 1).$$

Înmulțind aceste relații, deducem

$$m_{2r} = 1.3\ldots(2r-3)(2r-1) = \frac{(2r)!}{2^r \cdot r!}. \qquad \text{(VII, 2.21)}$$

Momentele centrate ale distribuției $N(m, \sigma)$ se pot calcula imediat dacă se cunosc momentele distribuției normale $N(0,1)$. Vom scrie

$$m_{2r+1} = \frac{1}{\sqrt{2\pi}\,\sigma} \int_{-\infty}^{+\infty} (x-m)^{2r+1} e^{-\frac{(x-m)^2}{2\sigma^2}} dx = 0,$$

deoarece funcția $(x-m)^{2r+1} e^{-\frac{(x-m)^2}{2\sigma^2}}$ este, de asemenea, impară și

$$m_{2r} = \frac{1}{\sqrt{2\pi}\,\sigma} \int_{-\infty}^{+\infty} (x-m)^{2r} e^{-\frac{(x-m)^2}{2\sigma^2}} dx$$

de unde făcînd substituția $y = \frac{x-m}{\sigma}$ și ținînd seama de (VII, 2.21) obținem

$$m_{2r} = \frac{1}{\sqrt{2\pi}\,\sigma} \int_{-\infty}^{+\infty} (x-m)^{2r} e^{-\frac{(x-m)^2}{2\sigma^2}} dx = \frac{\sigma^{2r}}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} y^{2r} e^{-\frac{y^2}{2}} dy =$$

$$= 1.3\ldots(2r-1)\,\sigma^{2r} = \frac{(2r)!}{2^r \cdot r!}\,\sigma^{2r}.$$

Oblicitatea și excesul *) pentru distribuția normală $N(m, \sigma)$ rezultă imediat:

$$\gamma_1 = \frac{m_3}{m_2^{\frac{3}{2}}} = 0, \quad \gamma_2 = \frac{m_4}{m_2^2} - 3 = \frac{3\sigma^4}{\sigma^4} - 3 = 0.$$

Prin urmare, oblicitatea și excesul pentru legea normală cea mai generală sînt egale cu zero.

De aici, cunoscutul procedeu al statisticii descriptive de a considera oblicitatea și excesul ca niște criterii pentru stabilirea normalității. Dacă aceste valori tipice sînt apropiate de zero, atunci este un indiciu că distribuția diferă puțin de distribuția normală.

După cum se observă, faptul că oblicitatea și excesul sînt egale cu zero constituie o condiție necesară, dar nu suficientă, și numai în acest sens criteriul de mai sus trebuie aplicat în practică.

B. *Funcția caracteristică a legii normale* $N(m, \sigma)$ *este* $e^{mti - \frac{\sigma^2 t^2}{2}}$.

Ne vom ocupa, întocmai ca și la calculul momentelor, în primul rînd, de determinarea funcției caracteristice a distribuției normale $N(0,1)$ și în al doilea rînd, de aceea a distribuției normale $N(m, \sigma)$. Astfel, funcția caracteristică a distribuției normale $N(0,1)$ este

$$\varphi(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{xti}\, e^{-\frac{x^2}{2}}\, dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} \sum_{n=0}^{\infty} \frac{t^n i^n}{n!} x^n\, e^{-\frac{x^2}{2}}\, dx.$$

Seria de sub semnul integral fiind uniform convergentă, vom putea să scrim mai departe

$$\varphi(t) = \sum_{n=0}^{\infty} \frac{t^n i^n}{n!} \cdot \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} x^n e^{-\frac{x^2}{2}}\, dx = \sum_{r=0}^{\infty} \frac{t^{2r}(-1)^r}{(2r)!} \cdot \frac{(2r)!}{2^r \cdot r!} =$$

$$= \sum_{r=0}^{\infty} (-1)^r \frac{\left(\frac{t^2}{2}\right)^r}{r!} = e^{-\frac{t^2}{2}}. \qquad \text{(VII,2,22)}$$

dacă $n = 2r$ este un număr par și avem în vedere relația (VII,2.21). Funcția caracteristică a distribuției normale $N(m, \sigma)$ rezultă imediat, observînd că în baza substituției $y = \frac{x - m}{\sigma}$, deducem

$$x = m + \sigma y.$$

Atunci, folosind expresia funcției caracteristice a distribuției normale $N(0,1)$, vom avea

$$\psi(t) = M(e^{xti}) = M[e^{(m+\sigma y)ti}] = e^{mti} M[e^{y(\sigma t)i}] = e^{mti} \varphi(\sigma t) = e^{mti - \frac{\sigma^2 t^2}{2}}. \qquad \text{(VII,2.23)}$$

C. *Dacă variabilele aleatoare independente* X_1 *și* X_2 *urmează legile normale* $N(m_1, \sigma_1)$ *și* $N(m_2, \sigma_2)$, *suma lor* $X_1 + X_2$ *urmează legea normală* $N\left(m_1 + m_2, \sqrt{\sigma_1^2 + \sigma_2^2}\right)$.

Într-adevăr, din teorema funcției caracteristice a sumei variabilelor independente rezultă

$$\psi(t) = e^{m_1 ti - \frac{\sigma_1^2 t^2}{2}}\, e^{m_2 ti - \frac{\sigma_2^2 t^2}{2}} = e^{(m_1 + m_2)ti - \frac{(\sigma_1^2 + \sigma_2^2)t^2}{2}}$$

care este funcția caracteristică a legii normale $N(m, \sigma)$ cu $m = m_1 + m_2$ și $\sigma^2 = \sigma_1^2 + \sigma_2^2$ în virtutea relației (VII,2.23).

*) Acești indici vor fi definiți la capitolul X.

Deci teorema este demonstrată.

D. *Mai general, dacă variabilele independente* $X_1, X_2, \ldots, X_n$ *urmează respectiv legile normale* $N(m_1, \sigma_1)$, $N(m_2, \sigma_2), \ldots N(m_n, \sigma_n)$, *atunci suma lor* $X_1 + X_2 + \ldots + X_n$ *urmează legea normală* $N\left(m_1 + m_2 + \ldots + m_n \,.\, \sqrt{\sigma_1^2 + \sigma_2^2 + \ldots + \sigma_n^2}\right)$.

Demonstrația este aceeași ca în cazul $n = 2$.

E. *Fie* $X_1, X_2, \ldots, X_n$ *un șir de variabile normale* $N(m_1, \sigma_1), \ldots, N(m_n, \sigma_n)$ *independente între ele. Variabila aleatoare* $a_1X_1 + a_2X_2 + \ldots + a_nX_n$ *urmează legea normală*

$$N\left(a_1m_1 + \ldots + a_nm_n \,.\, \sqrt{a_1^2\sigma_1^2 + \ldots + a_n^2\sigma_n^2}\right).$$

Această teoremă importantă ne arată că orice combinație liniară de legi normale este de asemenea o lege normală.

Vom demonstra mai întîi că dacă legea de probabilitate a lui X este o lege normală $N(m, \sigma)$, aX urmează legea normală $N(am, a\sigma)$.

Într-adevăr, am arătat că funcția caracteristică a legii normale $N(m, \sigma)$ este

$$\varphi(l) = M(e^{Xli}) = e^{mli - \frac{\sigma^2 l^2}{2}}.$$

Funcția caracteristică a variabilei aX prin definiție este

$$M(e^{aXli}) = e^{amli - \frac{a^2\sigma^2 l^2}{2}}$$

obținută din relația precedentă punînd al în locul lui t, adică tot funcția caracteristică a unei legi normale. Variabila a_kX_k, $(k = 1,2,\ldots, n)$, urmează atunci respectiv legea normală $N(a_km_k, a_k\sigma_k)$, $(k = 1, 2, \ldots, n)$. Nu avem decît să aplicăm variabilelor a_kX_k, $(k = 1, \ldots, n)$, teorema D, pentru ca să obținem proprietatea E.

Să facem acum următoarele particularizări: $a_k = \frac{1}{n}$, $m_k = m$, $\sigma_k = \sigma$, $(k = 1, 2, \ldots, n)$.

Rezultă astfel următoarea proprietate importantă pentru aplicație:

Dacă variabilele independente $X_1, X_2, \ldots, X_n$ *urmează legea normală* $N(m, \sigma)$, *media lor aritmetică* $\frac{X_1 + X_2 + \ldots + X_n}{n}$ *urmează legea normală* $N\left(m; \frac{\sigma}{\sqrt{n}}\right)$.

Această proprietate ne arată de ce atunci cînd facem n măsurători a unei lungimi, media lor aritmetică este mai aproape de valoarea reală a lungimii. Admițînd că erorile măsurătorii se distribuie după aceeași lege normală, media aritmetică a măsurătorilor se distribuie, după teorema de mai sus, tot după o lege normală cu aceeași medie, însă cu dispersia mult mai mică.

F. *Dacă* X_1 *și* X_2 *sînt două variabile independente și suma* $X_1 + X_2$ *urmează o lege normală, atunci fiecare dintre ele urmează o lege normală.*

Această interesantă proprietate a legii normale a fost găsită de H. Cramér. Avînd mai puțină importanță în aplicație, ne mărginim numai la enunțarea ei.

3. Legea polinomială (multinomială)

În capitolul III am văzut modul în care putem determina probabilitățile $P_{n(\lambda_1, \lambda_2, \ldots, \lambda_s)}$ în cazul urnei Bernoulli cu mai multe stări.

Vom face aici cîteva precizări suplimentare, cunoscînd că pentru această urnă se pun, în general, aceleași probleme ca și în cazul clasic al urnei Bernoulli, cînd avem două stări.

În cazul în care în urnă avem bile de s culori, probabilitatea $P_{n(\lambda_1, \lambda_2, \ldots, \lambda_s)}$ ca în n extracții să scoatem λ_1 bile de culoarea întîi, λ_2 bile de culoarea a doua, ... λ_s bile de culoarea s, este

$$P_{n(\lambda_1, \lambda_2, \ldots, \lambda_s)} = \frac{n!}{\lambda_1! \, \lambda_2! \ldots \lambda_s!} p_1^{\lambda_1} p_2^{\lambda_2} \ldots p_s^{\lambda_s}, \qquad \text{(VII,3.1)}$$

unde p_i este probabilitatea extracției unei bile de culoare i.

Formula (VII,3.1) exprimă *legea polinomială*.

Pentru valori mari ale lui n, obținem o valoare asimptotică a probabilității $P_{n(\lambda_1, \lambda_2, \ldots, \lambda_s)}$.

Într-adevăr, dacă introducem abaterile reduse

$$\xi_i = \frac{\lambda_i - np_i}{\sqrt{np_iq_i}},$$

și ne propunem să arătăm că pentru abaterile reduse îndeplinind condițiile $a_i \leqslant \xi_i < b_i$, unde a_i și b_i sînt constante, atunci avem următoarea convergență

$$\sqrt{n^{s-1}} \, P_{n(\lambda_1, \ldots, \lambda_s)} : \frac{e^{-\frac{1}{2}\sum q_i \xi_i^2}}{(2\pi)^{\frac{s-1}{2}} \sqrt{p_1 p_2 \ldots p_s}} \to s.$$

Putem da acestei teoreme o demonstrație elementară, bazîndu-ne pe formula lui Stirling

$$n! = \sqrt{2\pi n} \, n^n e^{-n} e^{\theta_n},$$

unde

$$|\theta_n| \leqslant \frac{1}{12n}.$$

Avem

$$P_{n(\lambda_1, \ldots, \lambda_s)} = \frac{n!}{\lambda_1! \ldots \lambda_s!} p_1^{\lambda_1} \ldots p_s^{\lambda_s}.$$

Aplicînd formula lui Stirling, rezultă

$$P_{n(\lambda_1, \ldots, \lambda_s)} = \frac{\sqrt{2\pi n} \, n^n e^{-n} e^{\theta_n - \sum_1^s \theta_{\lambda_i}}}{(2\pi)^{\frac{s}{2}} \prod_{i=1}^{s} \lambda_i^{\lambda_i + \frac{1}{2}} e^{-\lambda_i}} \prod_{i=1}^{s} p_i^{\lambda_i} = \frac{1}{(2\pi)^{\frac{s-1}{2}}} \sqrt{\frac{n}{\lambda_1 \ldots \lambda_s}} e^{\theta} \prod_{i=1}^{s} \left(\frac{np_i}{\lambda_i}\right)^{\lambda_i} =$$

$$= \frac{e^{\theta}}{(2\pi)^{\frac{s-1}{2}} \sqrt{p_1 \ldots p_s}} \prod_{i=1}^{s} \left(1 + \xi_i \sqrt{\frac{q_i}{np_i}}\right)^{-np_i - \xi_i \sqrt{np_i q_i} - \frac{1}{2}}$$

$$\theta = \theta_n - \theta_{\lambda_1} - \ldots - \theta_{\lambda_s}); \quad \sum_{i=1}^{s} \lambda_i = n.$$

Însă

$$\lg \prod_1^s \left(1+\xi_i \sqrt{\frac{q_i}{np_i}}\right)^{-\left(np_i+\xi_i\sqrt{npq_i}+\frac{1}{2}\right)} = -\sum_1^s \left(np_i+\xi_i\sqrt{npq_i}+\frac{1}{2}\right)\lg\left(1+\xi_i\sqrt{\frac{q_i}{np_i}}\right) =$$

$$= -\sum \left(np_i+\xi_i\sqrt{npq_i}+\frac{1}{2}\right)\left[\xi_i\sqrt{\frac{q_i}{np_i}}-\frac{1}{2}\frac{q_i\xi_i^2}{np_i}+\theta\left(\frac{1}{n^{\frac{3}{2}}}\right)\right] = -\frac{1}{2}\sum_{i=1}^s q_i\xi_i^2+\theta\left(\frac{1}{\sqrt{n}}\right)$$

deoarece

$$\sum \xi_i \sqrt{npq_i} = \sum (\lambda_i - np_i) = \sum \lambda_i - n\sum p_i = n-n = 0,$$

iar

$$|\theta| < \frac{1}{12}\left(\frac{1}{n}+\frac{1}{\lambda_1}+\ldots+\frac{1}{\lambda_s}\right) < \frac{1}{12n}\left(1+\sum \frac{1}{p_i\left(1+a_i\sqrt{\frac{q_i}{np_i}}\right)}\right).$$

Teorema este deci demonstrată.

4. Legea hipergeometrică

În aplicațiile practice ale calculului probabilităților se ivesc probleme care nu pot fi încadrate în schema probabilistică a urnei lui J. Bernoull și în consecință, rezolvarea lor cere în mod necesar cunoașterea altor legi de probabilitate. Una din aceste legi este și *legea hipergeometrică*.

Schema generală a acestei legi a fost arătată în capitolul III, unde prin formula

$$P = \frac{C_A^\alpha \cdot C_B^\beta}{C_{A+B}^n}, \quad \alpha+\beta = n$$

am determinat probabilitatea ca în n probe succesive să obținem α bile albe și β bile negre, dintr-o urnă în care se găsesc A bile albe și B bile negre, fără ca după fiecare extracție bila extrasă să se pună înapoi în urnă (schema bilei nerevenite).

Este vorba deci de n probe dependente, adică de un șir de variabile aleatoare *dependente*, unde fiecare dintre ele poate lua valorile $\xi_{k1}(1,0)$ sau $\xi_{k2}(0,1)$ cu probabilitățile respective P_{k1} și P_{k2}, $(k=1, 2,\ldots,n)$, $(P_{k1} + P_{k2} = 1)$, care urmează să fie determinate.

Tot în capitolul III am considerat și cazul general cînd într-o urnă se găsesc N bile de k culori și am determinat probabilitatea dată cu ajutorul formulei

$$P = \frac{C_{N_1}^{n_1}\cdot C_{N_2}^{n_2}\cdots C_{N_k}^{n_k}}{C_{N_1+N_2\ldots+N_k}^{n_1+n_2\ldots+n_k}}$$

ca în n extracții să avem $n_1, n_2, \ldots, n_k$ bile de culorile respective.

Vom face acum unele precizări în legătură cu proprietățile legii hipergeometrice :

$$P_{n(\alpha)} = \frac{C_A^{\alpha} \cdot C_B^{\beta}}{C_{A+B}^{\alpha+\beta}}. \qquad \text{(VII,4.1)}$$

1°. Probabilitatea ca în n extracții succesive să obținem oricare din valorile pe care le poate lua bila albă este evident egală cu 1 :

$$\sum_{\alpha=a}^{b} P_{n(\alpha)} = 1, \quad a = \max\,[0,\, n-B]\; b = \min(n;\; A)$$

2°. Pentru ca în n extracții să nu obținem nici o bilă albă avem probabilitatea

$$P_n(0) = \frac{B\,(B-1)(B-2)\ldots(B-n+1)}{(A+B)\,(A+B-1)\ldots(A+B-n+1)}, \quad n \leqslant B. \qquad \text{(VII,4.2)}$$

Folosind această relație, expresia (VII, 4.1) se poate pune sub forma

$$P_{n(\alpha)} = \frac{n(n-1)\ldots(n-\alpha+1)}{\alpha\,!} \cdot \frac{A(A-1)\ldots(A-\alpha+1)}{(B-n+1)(B-n+2)\ldots(B-\beta)} \cdot P_n(0). \text{(VII,4.3)}$$

Considerînd acum funcția generatoare

$$\Phi(t) = \sum_{\alpha=0}^{n} P_{n(\alpha)}\, t^{\alpha}$$

și notînd $n = -a$, $A = -b$, $B-n+1 = c$, rezultă, ținînd seama de (VII,4.2) :

$$\Phi(t) = P_n(0)\left[1 + \frac{abt}{1\cdot c} + \frac{a\,(a+1)}{1\cdot 2} \cdot \frac{b\,(b+1)}{c\,(c+1)}\, t^2 + \ldots\right]$$

adică seria hipergeometrică a lui Gauss, de la care vine și numele legii de probabilitate.

3°. Pentru valori mari ale lui $A + B$ se poate arăta că valoarea probabilității $P_{n(\alpha)}$ după legea hipergeometrică tinde către valoarea aceleiași probabilități după legea binomială, iar la limită tinde către legea normală.

Acest rezultat este firesc, deoarece cu cît numărul inițial de bile din urnă este mai mare, cu atît scoaterea unei bile din urnă exercită o acțiune mai mică asupra compoziției urnei ; la limită, se ajunge chiar la o succesiune de variabile independente.

Momente

Am reținut din cele de mai sus că în cazul legii de probabilitate hipergeometrică avem o succesiune de variabile aleatoare

$$x_1,\; x_2, \ldots,\; x_n$$

fiecare dintre ele luînd valorile $\xi_{i_1}(1,0)$ și $\xi_{i_2}(0,1)$ cu probabilitățile respective

$$p_{i_1} \text{ și } p_{i_2} \ (p_{i_1} + p_{i_2} = 1), \ (i = 1,2,\ldots, n).$$

Într-o singură probă, variabila aleatoare x_i poate să apară sau nu de unde rezultă legea de probabilitate

$$x_i \begin{bmatrix} 1 & 0 \\ p_{i_1} & p_{i_2} \end{bmatrix}.$$

Să calculăm acum primele momente obișnuite și momente centrate ale variabilei noastre

$$X = x_1 + x_2 + \ldots + x_n.$$

A. MOMENTE OBIȘNUITE. 1°. *Momentul obișnuit de ordinul întîi.* Prin definiție avem

$$M_1 = M(X) = \sum_{i=1}^{n} M(x_i) = \sum_{i=1}^{N} p_i. \qquad \text{(VII,4.4)}$$

Se demonstrează că $M_1 = np$, ca și în cazul bilei întoarse.

Pentru determinarea probabilității $P(x_i = 1)$ ca în extracția de rangul i să obținem o bilă albă, vom judeca după cum urmează:

Să presupunem că am numerotat cele N bile din urnă de la 1 pînă la A pe cele albe și de la $A+1$ pînă la N pe cele negre și le-am aranjat în ordinea crescătoare a indicilor

$$\underbrace{1,2\ldots, A\text{-}1, A,}_{\text{bile albe}} \qquad \underbrace{A+1,\ldots, N}_{\text{bile negre}}$$

și că rezultatele extracțiilor succesive s-au prezentat astfel:

$$\xi_1, \xi_2, \ldots \xi_i, \ldots, \xi_j, \ldots, \xi_N$$

unde, evident, numerele ξ_i, $(i = 1, 2, \ldots, N)$, reprezintă numerele $1, 2, \ldots, N$ luate cîte o singură dată, într-o ordine oarecare.

Observăm că numărul cazurilor posibile este egal cu $N!$, deoarece fiecare dintre aceste bile poate să apară în extracția de rangul i, $(i = 1, 2, \ldots, N)$.

Mai observăm apoi că numărul cazurilor favorabile ca în extracția i să ne apară o bilă albă este egală cu $A(N-1)!$. Într-adevăr, o bilă albă apărută în extracția de rangul i se poate asocia cu oricare dintre grupele de $N-1$ bile, apărute în extracțiile de celelalte ranguri, diferite de i; numărul acestor grupe este $(N-1)!$ și cum fiecare din cele A bile albe care pot ieși în extracția de rangul i se poate asocia la rîndul ei cu cele $(N-1)!$ grupe, rezultă că numărul cazurilor favorabile este tocmai $A(N-1)!$. Așadar, probabilitatea căutată este

$$P(x_i = 1) = \frac{A(N-1)!}{N!} = \frac{A}{N} = p.$$

Se observă că această probabilitate este independentă de i. Rezultă că expresia (VII,4.4) se poate scrie $M_1 = np$.

2°. *Momentul obişnuit de ordinul al doilea.* Prin definiţie acest moment are expresia

$$M_2 = M(X^2) = M\left(\sum_{i=1}^{n} x_i\right)^2 = \sum_{i=1}^{n} M(x_i^2) + 2\sum_{i\neq j} M(x_i x_j) = \\ = np + n(n-1)M(x_i x_j). \qquad \text{(VII,4.5)}$$

Ţinînd seama că probabilitatea $P(x_i = 1,\ x_j = 1)$ este egală cu $\frac{p(Np-1)}{N-1}$, relaţia (VII,4.5) devine

$$M_2 = np + n(n-1)\frac{(Np-1)p}{N-1} = \frac{Np-1}{N-1}pn^2 + \frac{N}{N-1}pqn = \\ = [(Np-1)\ pn^2 + Npqn]\ \frac{1}{N-1}, \text{ în care } q = 1-p. \qquad \text{(VII,4.6)}$$

Pentru determinarea probabilităţii $P(x_i = 1, x_j = 1)$, vom raţiona în felul următor:

Pentru calculul lui $M(x_i, x_j)$ va trebui să considerăm variabila $x_i x_j$ care poate lua valorile

$$(x_i = 1,\ x_j = 1);\ (x_i = 1,\ x_j = 0);\ (x_i = 0,\ x_j = 1);\ (x_i = 0,\ x_j = 0).$$

Produsul $x_i\ x_j$ ia singura valoare 1 diferită de zero cu probabilitatea $P\ (x_i = 1,\ x_j = 1)$ pe care trebuie s-o determinăm.

Numărul cazurilor egal posibile este $N!$, deoarece în fiecare din extracţiile de rangul i şi j poate să iasă oricare din cele n bile din urnă.

Numărul cazurilor favorabile se obţine asociind în toate modurile posibile cele două bile ieşite în extracţiile de rangul i şi j cu fiecare dintre cele $(N-2)!$ grupe ce se pot forma cu restul de $N-2$ bile apărute în celelalte extracţii. În extracţiile de rangul i şi j pot apare însă oricare dintre cele A bile albe ceea ce înseamnă că numărul cazurilor favorabile este egal cu $A(A-1)\ (N-2)!$

Prin urmare, probabilitatea căutată este

$$P(x_i = 1,\ x_j = 1) = \frac{A(A-1)(N-2)!}{N!} = \frac{A(A-1)}{N(N-1)} = \frac{p(Np-1)}{(N-1)}.$$

După cum se vede, această probabilitate nu depinde de i şi j.

B. MOMENTE CENTRATE. A determina momentele centrate ale unei variabile aleatoare ce se supune legii hipergeometrice înseamnă a considera succesiunea de variabile aleatoare

$$y_i \begin{bmatrix} 1-p & -p \\ p_{i_1} & p_{i_2} \end{bmatrix}, \qquad (i = 1,2,\ldots,n)$$

în care prin y_i am notat abaterea de la media p a variabilei X, şi a calcula momentele sumei

$$Y = y_1 + y_2 + \ldots + y_n.$$

1°. *Momentul centrat de ordinul întîi*:

$$M_1 = M(Y) = \sum_{i=1}^{n} M(y_i) = 0.$$

2°. *Momentul centrat de ordinul al doilea.* Pornind de la formula generală a momentului centrat de ordinul al doilea

$$M_2 = M(Y^2) = M\left(\sum_{i=1}^{n} y_i\right)^2 = \sum_{i=1}^{n} M(y_i^2) + 2\sum_{i\neq j}(y_i y_j) = \\ = npq + n(n-1)\,M(y_i y_j). \qquad \text{(VII,4.7)}$$

și determinînd apoi media variabilei aleatoare $M(y_i\, y_j)$ pe care o introducem în (VII,4.7) se obține următoarea valoare a lui $M_2 = \sigma^2$:

$$\sigma^2 = M_2 = \frac{N-n}{N-1} \cdot npq. \qquad \text{(VII,4.8)}$$

După cum se poate vedea, dispersia σ^2 corespunzătoare la n probe în legea hipergeometrică diferă de dispersia de la urna lui Bernoulli prin factorul $\frac{N-n}{N-1}$, care este întotdeauna mai mic decît unu.

În mod analog putem obține și celelalte momente centrate de ordin superior lui doi.

Mai trebuie să observăm că

$$\lim_{N\to\infty} M_2 = \lim_{N\to\infty} \frac{N-n}{N-1}\, npq = npq$$

ceea ce înseamnă că la limită pentru $N\to\infty$, momentul centrat de ordinul al doilea al legii de probabilitate hipergeometrice tinde către momentul centrat de ordinul al doilea al legii binomiale.

5. Legea Poisson (legea evenimentelor rare)

Știm că legea binomială de probabilitate, exprimată prin formula

$$P_{n(\alpha)} = C_n^{\alpha} p^{\alpha} q^{n-\alpha} \qquad \text{(VII,5.1)}$$

ne arată valoarea probabilității ca un eveniment E cu probabilitatea constantă p de realizare, în n probe succesive independente, să se verifice de α ori.

Mai știm că, pentru valori mari ale lui n, probabilitatea $P_{n(\alpha)}$ poate fi pusă, folosind formula lui Moivre-Stirling, sub o altă formă, apropiată de tipul legii normale de probabilitate.

Aplicarea formulei lui Moivre-Stirling este însă legată de anumite condiții care nu sînt întotdeauna îndeplinite. Dacă p este foarte mic iar n destul de mare ca să avem

$$np = \lambda, \qquad \text{(VII,5.2)}$$

în care λ este o constantă, se poate ușor demonstra că valoarea asimptotică a probabilității (VII,5.2) este de altă formă și anume:

$$C_n^\alpha p^\alpha q^{n-\alpha} \approx \frac{\lambda^\alpha}{\alpha!} e^{-\lambda}. \qquad \text{(VII,5.3)}$$

Într-adevăr, avem

$$P_{n(\alpha)} = \frac{n(n-1)\dots(n-\alpha+1)}{\alpha!} p^\alpha q^{n-\alpha}.$$

Înlocuind probabilitatea p prin valoarea rezultată din (VII,5.2), obținem

$$P_{n(\alpha)} = \frac{n(n-1)\dots(n-\alpha+1)}{\alpha!} \cdot \left(\frac{\lambda}{n}\right)^\alpha \cdot \left(1 - \frac{\lambda}{n}\right)^{n-\alpha}.$$

Cum însă

$$\lim_{n-\infty} \frac{n(n-1)\dots(n-\alpha+1)}{n^\alpha} = 1 \text{ și } \lim_{n\to\infty} \left(1 - \frac{\lambda}{n}\right)^{n-\alpha} = e^{-\lambda},$$

putem scrie valoarea asimptotică a probabilității

$$P_{n(\alpha)} \approx \frac{\lambda^\alpha}{\alpha!} e^{-\lambda}, \qquad (\alpha = 0,1,\dots,\infty).$$

Formula (VII,5.3) a fost găsită pentru prima oară de Poisson și poartă numele de *legea de probabilitate a lui Poisson*. Această lege se întîlnește în cazul evenimentelor care se întîmplă foarte rar. Din această cauză ea mai este cunoscută și sub denumirea de *legea evenimentelor rare*.

L. Bortkievici a fost cel dintîi care s-a ocupat de distribuțiile statistice care urmează legea de probabilitate a evenimentelor rare.

Pentru legea Poisson avem evident

$$\sum_{\alpha=0}^{\infty} \frac{\lambda^\alpha}{\alpha!} e^{-\lambda} = e^{-\lambda} \sum_{\alpha=0}^{\infty} \frac{\lambda^\alpha}{\alpha!} = 1.$$

MOMENTE. 1°. *Momente obișnuite*. Se demonstrează că pentru primele patru momente obișnuite avem următoarele valori:

$$M_1 = \lambda,$$
$$M_2 = \lambda + \lambda^2,$$
$$M_3 = \lambda + 3\lambda^2 + \lambda^3,$$
$$M_4 = \lambda + 7\lambda^2 + 6\lambda^3 + \lambda^4.$$

Scanned with OKEN Scanner

2°. *Momente centrate.* Aceste momente se deduc ușor din momentele obișnuite. Rezultă următoarele valori pentru primele patru momente centrate:

$$m_1 = 0,$$
$$m_2 = M_2 - M_1^2 = \lambda + \lambda^2 - \lambda^2 = \lambda,$$
$$m_3 = M_3 - 3M_2M_1 + 2M_1^3 = \lambda + 3\lambda^2 + \lambda^3 - 3(\lambda + \lambda^2)\lambda + 2\lambda^3 = \lambda$$
$$m_4 = M_4 - 4M_3M_1 + 6M_2M_1^2 - 3M_1^4 = \lambda + 3\lambda^2.$$

O constatare interesantă se poate face și anume: momentele m_1, m_2 și m_3 ale legii Poisson sînt egale, valoarea lor comună fiind $\frac{\lambda^\alpha}{\alpha!}\lambda$.

Ținînd seama de formula (VII,5.3), avem

$$M(\alpha) = \sum_1^\infty \alpha \cdot \frac{\lambda^\alpha}{\alpha!} e^{-\lambda} = \lambda e^{-\lambda} \cdot \sum_1^\infty \frac{\lambda^{\alpha-1}}{(\alpha-1)!} = \lambda$$

deoarece

$$\sum_1^\infty \frac{\lambda^{\alpha-1}}{(\alpha-1)!} = e^\lambda.$$

În mod analog:

$$M[\alpha(\alpha-1)] = \lambda^2,$$
$$M[\alpha(\alpha-1)(\alpha-2)] = \lambda^3,$$
$$M[\alpha(\alpha-1)(\alpha-2)(\alpha-3)] = \lambda^4.$$

Din dezvoltarea acestor relații, ținînd seama că valoarea medie a unei sume este egală cu suma valorilor medii ale componentelor ei, obținem:

$$M(\alpha) = \lambda,$$
$$M(\alpha^2) - M(\alpha) = \lambda^2.$$
$$M(\alpha^3) - 3M(\alpha^2) + 2M(\alpha) = \lambda^3,$$
$$M(\alpha^4) - 6M(\alpha^3) + 11M(\alpha^2) - 6M(\alpha) = \lambda^4.$$

Avem un sistem de patru ecuații liniare cu patru necunoscute: $M(\alpha)$, $M(\alpha^2)$, $M(\alpha^3)$, $M(\alpha^4)$, din rezolvarea cărora rezultă momentele obișnuite pe care le-am scris mai sus.

Funcția caracteristică a legii Poisson rezultă imediat din definiția ei:

$$\Phi(t, \lambda) = M(e^{\alpha ti}) = \sum_{\alpha=0}^\infty e^{\alpha ti} \frac{\lambda^\alpha}{\alpha!} e^{-\lambda} = e^{\lambda(e^{ti}-1)}.$$

Din această expresie a funcției caracteristice rezultă două teoreme importante:

TEOREMĂ. *Dacă X și Y sînt două variabile aleatoare independente, fiecare din ele supunîndu-se legii de probabilitate Poisson cu parametrii respectivi λ și μ, atunci suma lor $X + Y$ urmează, de asemenea, o lege Poisson cu parametrul $(\lambda + \mu)$.*

Avem

$$P(X = \alpha) = \frac{\lambda^\alpha}{\alpha!} e^{-\lambda}; \quad P(Y = \alpha) = \frac{\mu^\alpha}{\alpha!} e^{-\mu}, \qquad (\alpha = 0,1,\ldots,\infty).$$

Atunci funcția caracteristică a variabilei $X + Y$, în baza proprietății funcției caracteristice a sumelor de variabile independente, este

$$\Phi(t,\lambda).\ \Phi(t,\mu) = e^{\lambda(e^{ti}-1)} \cdot e^{\mu(e^{ti}-1)} = e^{(\lambda+\mu)(e^{ti}-1)} = \Phi(t,\lambda+\mu).$$

Această teoremă pune în evidență o proprietate remarcabilă a legii Poisson și anume că este *o lege stabilă*.

TEOREMĂ. *Dacă parametrul λ al unei legi Poisson crește necontenit, atunci variabila normală convenabil tinde către legea normală a lui Laplace.*

Să introducem abaterea redusă

$$\frac{\alpha-\lambda}{\sqrt{\lambda}}, \qquad (\alpha = 0,1,\ldots,\infty)$$

și să considerăm funcția caracteristică a acestei abateri. Vom scrie

$$\psi(t,\lambda) = M\left(e^{\frac{\alpha-\lambda}{\sqrt{\lambda}}ti}\right) = e^{-\sqrt{\lambda}ti} M\, e^{\left(\alpha\frac{t}{\sqrt{\lambda}}i\right)} = e^{-\sqrt{\lambda}ti}\varphi\left(\frac{t}{\sqrt{\lambda}},\lambda\right) =$$

$$= e^{-\sqrt{\lambda}ti} \cdot e^{\lambda\left(e^{\frac{t}{\sqrt{\lambda}}i}-1\right)} = e^{\lambda\left(e^{\frac{t}{\sqrt{\lambda}}i}-1\right)-\sqrt{\lambda}ti} =$$

$$= \exp\left\{\lambda\left(e^{\frac{t}{\sqrt{\lambda}}i}-1\right)-\sqrt{\lambda}ti\right\} = \exp\left\{\lambda\left[\frac{t}{\sqrt{\lambda}}i-\frac{t^2}{2\lambda}+\right.\right.$$

$$\left.\left.+\,0\left(\lambda^{-\frac{3}{2}}\right)\right]-\sqrt{\lambda}ti\right\} = \exp\left\{-\frac{t^2}{2}+0\left(\lambda^{-\frac{1}{2}}\right)\right\},$$

de unde deducem

$$\lim_{\lambda\to\infty}\psi(t,\lambda) = e^{-\frac{t^2}{2}}$$

care este funcția caracteristică a legii normale cu dispersia $\sigma^2 = 1$. Dacă funcțiile caracteristice $\psi(t,\lambda)$ tind către funcția caracteristică a legii normale, atunci se știe că funcțiile de repartiție corespunzătoare vor tinde de asemenea către funcția de repartiție a legii normale.

Teorema este deci demonstrată.

6. Legea de probabilitate χ^2

Să considerăm șirul de variabile aleatoare independente

$$x_1,\ x_2,\ \ldots,\ x_s,$$

unde fiecare dintre aceste variabile urmează legea normală de tipul $N(0,\sigma)$.

Notînd prin y suma pătratelor tuturor acestor variabile,

$$y = x_1^2 + x_2^2 + \ldots + x_s^2,$$

obținem o altă variabilă aleatoare — variabila y.

În teoria probabilităților se demonstrează că densitatea de probabilitate a variabilei y este

$$\frac{1}{2^{\frac{s}{2}}\sigma^s\Gamma\left(\frac{s}{2}\right)} y^{\frac{s}{2}-1} e^{-\frac{y}{2\sigma^2}}. \qquad \text{(VII,6.1)}$$

Ea poartă numele de legea de probabilitate χ^2 cu s grade de libertate, $(\chi^2 = y)$.

Funcția de frecvență a distribuției χ^2 a lui Helmert-Pearson se poate stabili pe baza unor considerații geometrice.

Să presupunem, astfel, că într-un spațiu euclidian cu s dimensiuni, se dă hipersfera cu centrul în origine și cu raza $R = y^{\frac{1}{2}}$:

$$x_1^2 + x_2^2 + \ldots + x_s^2 = y.$$

Funcția de frecvență $f(y)\,dy$ este, evident, egală cu probabilitatea ca un punct $M(x_1, x_2, \ldots, x_s)$ din acest spațiu cu s dimensiuni să cadă în domeniul (V) cuprins între două hipersfere cu centrele în origine și cu razele respectiv egale cu $y^{\frac{1}{2}}$ și $(y+dy)^{\frac{1}{2}}$.

Cunoscînd că probabilitatea elementară pentru punctul $M(x_1, x_2, \ldots, x_s)$ este

$$\frac{1}{(2\pi)^{\frac{s}{2}}\sigma^s} e^{-\frac{\sum_{i=1}^{s} x_i^2}{2\sigma^2}} dx_1\, dx_2 \ldots dx_s,$$

vom avea

$$f(y)dy = P\left(y < \sum_{i=1}^{s} x_i^2 \leqslant y + dy\right) =$$

$$= \frac{1}{(2\pi)^{\frac{s}{2}}\sigma^s} \iint\ldots\int_V e^{-\frac{\sum_{i=1}^{s} x_i^2}{2\sigma^2}} dx_1\, dx_2 \ldots dx_s. \qquad \text{(VII,6.2)}$$

Aplicînd formula mediei, expresia (VII,6.2) devine

$$f(y)dy = \frac{1}{(2\pi)^{\frac{s}{2}}\sigma^s} A \cdot V, \qquad \text{(VII,6.3)}$$

unde V este domeniul de integrare egal cu volumul cuprins între cele două hipersfere cu centrele în origine și cu razele respectiv egale cu $y^{\frac{1}{2}}$ și $(y + dy)^{\frac{1}{2}}$, iar A valoarea funcției $e^{-\frac{\Sigma x_i^2}{2\sigma^2}}$ într-un punct determinat din V.

Se știe că volumul unei hipersfere, într-un spațiu euclidian cu s dimensiuni, este dat de formula

$$\frac{\pi^{\frac{s}{2}}}{\Gamma\left(\frac{s}{2}+1\right)} \cdot R^s,$$

R fiind raza hipersferei.

Atunci, volumul cuprins între cele două hipersfere cu centrele în origine şi cu razele respectiv egale cu $y^{\frac{1}{2}}$ şi $(y + dy)^{\frac{1}{2}}$ este

$$V = \frac{\pi^{\frac{s}{2}}}{\Gamma\left(\frac{s}{2}+1\right)}\left[(y+dy)^{\frac{s}{2}} - y^{\frac{s}{2}}\right] = \frac{\pi^{\frac{s}{2}}\, s}{2\Gamma\left(\frac{s}{2}+1\right)} y^{\frac{s}{2}-1}\, dy = \frac{\pi^{\frac{s}{2}}}{\Gamma\left(\frac{s}{2}\right)} y^{\frac{s}{2}-1}\, dy \quad \text{(VII, 6.4)}$$

dacă neglijăm termenii dezvoltării binomului $(y + dy)^{\frac{s}{2}}$, care conţin infiniţi mici de ordin superior lui dy, şi înlocuim $2\Gamma\left(\frac{s}{2}+1\right)$ cu $s\Gamma\left(\frac{s}{2}\right)$.

Înlocuind în (VII,6.3) pe V cu expresia lui din (VII,6.4) şi luînd $A = e^{-\frac{y}{2\sigma^2}}$, deoarece $y < \sum_{i=1}^{s} x_i^2 \leqslant y + dy$, obţinem

$$f(y)dy = \frac{1}{(2\pi)^{\frac{s}{2}}\cdot\sigma^s} \cdot \frac{\pi^{\frac{s}{2}}}{\Gamma\left(\frac{s}{2}\right)} \cdot y^{\frac{s}{2}-1} \cdot e^{-\frac{y}{2\sigma^2}} \cdot dy =$$

$$= \frac{1}{2^{\frac{s}{2}}\cdot\sigma^s\cdot\Gamma\left(\frac{s}{2}\right)} \cdot y^{\frac{s}{2}-1}\, e^{-\frac{y}{2\sigma^2}}\, dy,$$

de unde rezultă, prin schimbarea de variabilă $y = x$, expresia funcţiei de frecvenţă a distribuţiei χ^2 a lui Helmert-Pearson

$$f(x) = \frac{1}{2^{\frac{s}{2}}\sigma^s\Gamma\left(\frac{s}{2}\right)} x^{\frac{s}{2}-1} \cdot e^{-\frac{x}{2\sigma^2}}. \quad \text{(VII, 6.1')}$$

Distribuţia cu această funcţie de frecvenţă o vom nota pe viitor cu $H(s, \sigma)$. Ea este o distribuţie cu s grade de libertate şi cu parametrul σ.

La expresia funcţiei de frecvenţă $f(x)$ se poate ajunge, de asemenea, determinînd legea de probabilitate a variabilei x_i^2.

Pentru aceasta, vom stabili mai întîi funcţia de repartiţie a acestei variabile.

Vom scrie deci

$$F(y) = P(x_i^2 \leqslant y) = P(-\sqrt{y} \leqslant x_i \leqslant \sqrt{y}) =$$

$$= \frac{1}{\sqrt{2\pi}\cdot\sigma}\int_{-\sqrt{v}}^{\sqrt{v}} e^{-\frac{x_i^2}{2\sigma^2}}\, dx_i = \frac{2}{\sqrt{2\pi}\cdot\sigma}\int_0^{\sqrt{v}} e^{-\frac{x_i^2}{2\sigma^2}}\, dx_i,$$

deoarece, fiecare dintre variabilele x_i, $(i = 1, 2, \ldots, s)$, urmează legea normală $N(0, \sigma)$, de unde rezultă

$$f(y) = \frac{dF(y)}{dy} = \frac{1}{\sqrt{2\pi}\cdot\sigma\cdot y^{\frac{1}{2}}} e^{-\frac{y}{2\sigma^2}},$$

adică legea de probabilitate a variabilei x_i^2.

Acum vom putea găsi funcția caracteristică a variabilei x_i^2, care, în baza definiției sale, se scrie

$$M(e^{vti}) = \frac{1}{\sqrt{2\pi}\cdot\sigma}\int_0^\infty y^{-\frac{1}{2}} e^{-\frac{y}{2\sigma^2}} e^{vti}\,dy = \frac{1}{\sqrt{2\pi}\,\sigma}\int_0^\infty y^{\frac{1}{2}-1} e^{-\frac{1-2\sigma^2 ti}{2\sigma^2}y}\,dy.$$

Din definiția funcției $\Gamma(x)$ se știe că

$$\int_0^\infty x^{p-1}\cdot e^{-ax}\cdot dx = \frac{\Gamma(p)}{a^p}.$$

Atunci, expresia funcției caracteristice, observînd că $p = \frac{1}{2}$ și $a = \frac{1-2\sigma^2 ti}{2\sigma^2}$, devine

$$M(e^{vti}) = \frac{1}{\sqrt{2\pi}\cdot\sigma}\,\frac{\Gamma\left(\frac{1}{2}\right)}{\left(\frac{(1-2\sigma^2 ti}{2\sigma^2}\right)^{\frac{1}{2}}} = \frac{1}{(1-2\sigma^2 ti)^{\frac{1}{2}}} = (1-2\sigma^2 ti)^{-\frac{1}{2}} \qquad \left|\,\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}\,\right|.$$

Aplicînd teorema funcției caracteristice a sumei de variabile independente, obținem imediat funcția caracteristică a variabilei $\sum_{i=1}^{s} x_i^2$:

$$\varphi(t) = (1-2\sigma^2 ti)^{-\frac{s}{2}}.$$

Pe de altă parte,

$$\frac{a^p}{\Gamma(p)}\int_0^\infty x^{p-1}\cdot e^{-ax}\,dx = 1,$$

ceea ce înseamnă că funcția

$$f(x) = \frac{a^p}{\Gamma(p)}\cdot x^{p-1} e^{-ax}$$

este funcție de frecvență a distribuției care conduce la funcția $\Gamma(p)$. Funcția caracteristică, corespunzătoare acestei legi, este

$$\Psi(t) = \frac{a^p}{\Gamma(p)}\int_0^\infty x^{p-1} e^{-ax} e^{xti}\,dx = \frac{a^p}{\Gamma(p)}\int_0^\infty x^{p-1} e^{-(a-ti)x}\,dx =$$

$$= \frac{a^p}{\Gamma(p)}\cdot\frac{\Gamma(p)}{(a-ti)^p} = \left(1-\frac{1}{a}ti\right)^{-p}.$$

În cazul nostru avem

$$p = \frac{s}{2} \text{ și } a = \frac{1}{2\sigma^2}$$

și regăsim funcția de frecvență a distribuției x^2 a lui Helmert-Pearson:

$$f(x) = \frac{\left(\frac{1}{2\sigma^2}\right)^{\frac{s}{2}}}{\Gamma\left(\frac{s}{2}\right)} x^{\frac{s}{2}-1} e^{-\frac{x}{2\sigma^2}} = \frac{1}{2^{\frac{s}{2}}\sigma^s\Gamma\left(\frac{s}{2}\right)} x^{\frac{s}{2}-1} e^{-\frac{x}{2\sigma^2}}.$$

Legea de probabilitate (VII,6.1′) se schimbă o dată cu schimbarea numărului gradelor de libertate. După cum se vede, în cazul legii specifice

de probabilitate $H(s, \sigma)$ se introduce o noțiune nouă a gradelor de libertate. Prin numărul gradelor de libertate se înțelege numărul variabilelor independente a căror variație nu suferă nici o restricție. De pildă, dacă considerăm o variabilă x care poate lua valori diferite $x_1, x_2, .., x_s$, fiecare cu probabilitățile $p_1, p_2, ..., p_s$ și avem, evident $p_1 + p_2 + ... + p_s = 1$, în acest caz avem o repartiție cu s valori componente și dacă cunoaștem $s-1$ dintre probabilități putem afla și ultima. Sîntem în cazul unei variabile cu $s-1$ grade de libertate.

Dacă notăm cu s numărul gradelor de libertate, cu s' numărul componentelor varabilei χ^2 și cu k numărul restricțiilor liniare care au loc la calculul lui χ^2, atunci

$$s = s' - k.$$

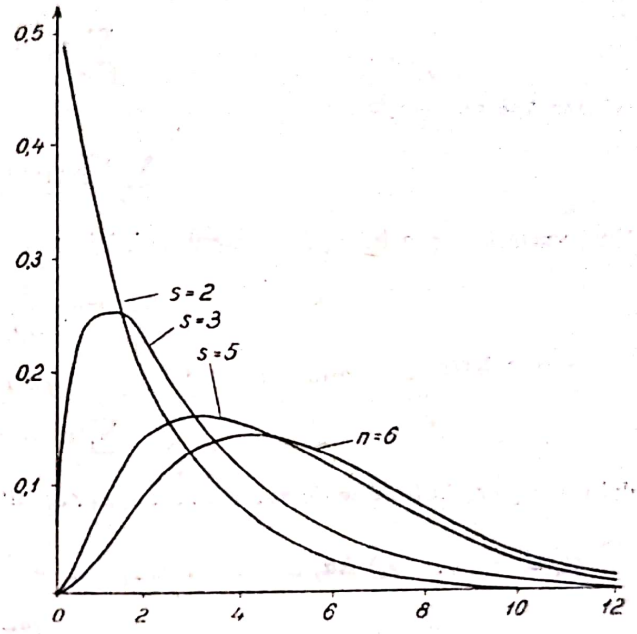


Fig. 9

Pentru legile de probabilitate $H(s, \sigma)$ cu un număr determinat de grade de libertate s-au construit tabele numerice (vezi tabelul anexă II).

În graficul din figura 9 sînt date legile de probabilitate $H(s, \sigma)$ cu $s = 2$, $s = 3$, $s = 5$, $s = 6$.

Se observă că în cazul cînd s crește, legea de probabilitate $H(s, \sigma)$ are tendința de a se apropia de legea normală.

Distribuția $H(s, \sigma)$ poate fi considerată și sub o altă formă. Să considerăm o variabilă χ care poate lua valorile $x_1, x_2, \ldots, x_s$, cu probabilitățile $p_1, p_2, \ldots, p_s$. Supunînd această variabilă la n probe independente și notînd prin α_1 de cîte ori a fost luată valoarea x_1, prin α_2 de cîte ori a fost luată valoarea x_2 etc., avem

$$\alpha_1 + \alpha_2 + \ldots + \alpha_s = n.$$

În acest caz χ^2 poate fi exprimat sub forma

$$\chi^2 = \frac{(\alpha_1 - np_1)^2}{np_1} + \frac{(\alpha_2 - np_2)^2}{np_2} + \ldots + \frac{(\alpha_s - np_s)^2}{np_s} \qquad \text{(VII,6.5)}$$

în care, evident, $\alpha_1, \alpha_2, \ldots, \alpha_s$ sînt valori empirice, iar $np_1, np_2, \ldots, np_s$ sînt valorile teoretice corespunzătoare. Egalitatea (VII, 6.5) are loc numai cînd n crește necontenit.

Am văzut din studiul distribuției multinomiale că valoarea asimptotică a probabilității $P_n(\alpha_1, \alpha_2, \ldots, \alpha_s)$ este dată de expresia

$$P_n(\alpha_1, \alpha_2, \ldots, \alpha_s) \simeq \frac{e^{-\frac{1}{2}\sum_{i=1}^{s} q_i \xi_i^2}}{(2\pi n)^{\frac{s-1}{2}} \cdot \sqrt{p_1 \cdot p_2 \cdots p_s}},$$

unde valorile $\alpha_1, \alpha_2, \ldots, \alpha_s$ intervin numai prin intermediul funcției

$$\chi^2 = \sum_{i=1}^{s} q_i \xi_i^2. \tag{VII,6.6}$$

Relația

$$\sum_{i=1}^{s} \alpha_i = n$$

echivalează cu relația

$$\sum_{i=1}^{s} \sqrt{p_i q_i} \cdot \xi_i = 0. \tag{VII,6.7}$$

Unei variații $\Delta\alpha_i$ a lui α_i îi corespunde lui ξ_i variația

$$\Delta\xi_i = \frac{\Delta\alpha_i}{\sqrt{np_iq_i}}.$$

Deoarece relația

$$\sum_{i=1}^{s} \alpha_i = n$$

definește pe α_s în funcție de $\alpha_1, \alpha_2, \ldots, \alpha_{s-1}$, rezultă că

$$P_n(\alpha_1, \alpha_2, \ldots. \alpha_s)\, \Delta\alpha_1\, \Delta\alpha_2 \ldots \Delta\alpha_{s-1} \simeq \frac{\sqrt{q_1 \cdot q_2 \ldots q_{s-1}}}{(2\pi)^{\frac{s-1}{2}} \sqrt{p_s}} \cdot e^{-\frac{\chi^2}{2}} \cdot \Delta\xi_1 \cdot \Delta\xi_2 \ldots \Delta\xi_{s-1}.$$

În locul distribuției discontinue, definită de $P_n(\alpha_1, \alpha_2, \ldots, \alpha_s)$, să considerăm distribuția continuă definită de probabilitatea elementară

$$dP = \frac{\sqrt{q_1 \cdot q_2 \ldots q_{s-1}}}{(2\pi)^{\frac{s-1}{2}} \sqrt{p_s}} \cdot e^{-\frac{\chi^2}{2}}\, d\xi_1 \cdot d\xi_2 \ldots d\xi_{s-1}$$

împreună cu relațiile (VII,6.4) și (VII,6.7).

Ne propunem să determinăm distribuția variabilei pozitive χ^2.

Substituția

$$\xi_i = \chi\eta_i \qquad (i = 1, 2, \ldots, s)$$

înlocuiește variabilele ξ_i cu variabilele $\eta_1, \eta_2, \ldots, \eta_s; \chi$ legate prin relațiile

$$\sum_{i=1}^{s} q_i \eta_i^2 = 1 \tag{VII,6.8}$$

$$\sum_{i=1}^{s} \sqrt{p_i q_i} \cdot \eta_i = 0 \tag{VII,6.9}$$

deduse din (VII,6.6) și (VII,6.7).

Considerînd $\eta_1, \eta_2, \ldots, \eta_{s-2}, \chi$ ca variabile independente, iar η_{s-1} și η_s ca definite cu ajutorul relațiilor (VII,6.8), (VII,6.9), rezultă că

$$q_h\eta_h + q_{s-1}\eta_{s-1}\frac{\partial\eta_{s-1}}{\partial\eta_h} + q_s\eta_s\frac{\partial\eta_s}{\partial\eta_h} = 0, \qquad (h = 1, 2, \ldots, s-2) \tag{VII,6.10}$$

$$\sqrt{p_hq_h} + \sqrt{p_{s-1}q_{s-1}} \cdot \frac{\partial\eta_{s-1}}{\partial\eta_h} + \sqrt{p_sq_s} \cdot \frac{\partial\eta_s}{\partial\eta_h} = 0. \tag{VII,6.11}$$

Introducînd notaţia

$$D_i = \begin{vmatrix} q_i\,\eta_i & q_s\,\eta_s \\ \sqrt{p_i q_i} & \sqrt{p_s q_s} \end{vmatrix},$$

din (VII,6.10) şi (VII,6.11) se obţine

$$\frac{\partial \eta_{s-1}}{\partial \eta_h} = -\frac{D_h}{D_{s-1}}.$$

Din determinantul funcţional

$$\frac{D(\xi_1, \xi_2, \ldots, \xi_{s-1})}{D(\eta_1, \eta_2, \ldots, \eta_{s-2}\ \chi)} = \begin{vmatrix} \chi & \ldots & 0 & \eta_1 \\ \ldots & \ldots & \ldots & \ldots \\ 0 & \ldots & \chi & \eta_{s-2} \\ \chi\dfrac{\partial \eta_{s-1}}{\partial \eta_1} & \ldots & \chi\dfrac{\partial \eta_{s-1}}{\partial \eta_{s-2}} & \eta_{s-1} \end{vmatrix} = \chi^{s-2} \begin{vmatrix} 1 & \ldots & 0 & \eta_1 \\ \ldots & \ldots & \ldots & \ldots \\ 0 & \ldots & 1 & \eta_{s-2} \\ \dfrac{\partial \eta_{s-1}}{\partial \eta_1} & \ldots & \dfrac{\partial \eta_{s-1}}{\partial \eta_{s-2}} & \eta_{s-1} \end{vmatrix}$$

scăzînd din ultima coloană celelalte coloane înmulţite cu $\eta_1, \ldots, \eta_{s-2}$ respectiv, deducem

$$\frac{D(\xi_1, \xi_2, \ldots, \xi_{s-1})}{D(\eta_1, \eta_2, \ldots, \eta_{s-2}, \chi)} = \chi^{s-2}\left(\eta_{s-1} - \sum_{h=1}^{s-2} \eta_h \frac{\partial \eta_{s-1}}{\partial \eta_h}\right) = \chi^{s-2}\,\frac{\sum_{i=1}^{s-1} \eta_i D_i}{D_{s-1}} =$$

$$= \chi^{s-2}\,\frac{\begin{vmatrix} \sum_{i=1}^{s-1} q_i\,\eta_i^2 & q_s\,\eta_s \\ \sum_{i=1}^{s-1} \sqrt{q_i\,p_i}\,\eta_i & \sqrt{p_s\,q_s} \end{vmatrix}}{D_{s-1}} = \chi^{s-2}\,\frac{\begin{vmatrix} 1 - q_s\eta_s^2 & q_s\eta_s \\ -\sqrt{p_s q_s}\,\eta_s & \sqrt{p_s q_s} \end{vmatrix}}{D_{s-1}} = \chi^{s-2}\,\frac{\sqrt{p_s\,q_s}}{D_{s-1}}.$$

Prin substituţia

$$\xi_i = \chi\eta_i, \qquad (i = 1, 2, \ldots, s),$$

fiecărui punct $(\xi_1, \xi_2, \ldots, \xi_{s-1})$ îi corespunde un singur punct $(\eta_1, \eta_2, \ldots, \eta_{s-2}, \chi)$, $(\chi > 0)$, obţinut prin determinarea lui ξ_s din (VII,6.7), a lui χ din (VII,6.6) şi a lui η_i din această substituţie.

Reciproc, fiecărui punct $(\eta_1, \eta_2, \ldots, \eta_{s-2}, \chi)$, $(\chi > 0)$, cu

$$\sum_{h=1}^{s-2} q_h\,\eta_h^2 < 1,$$

îi corespund două puncte $(\xi_1, \xi_2, \ldots, \xi_{s-1})$, obţinute prin determinarea a două sisteme de valori (η_{s-1}, η_s) din (VII,6.6), (VII,6.9) şi substituţia $\xi_i = \chi\eta_i$.

Deci, prin substituţia $\xi_i = \chi\eta_i$ spaţiul $\xi_1, \xi_2, \ldots, \xi_{s-1}$ acoperă de două ori spaţiul $\eta_1, \eta_2, \ldots, \eta_{s-1}$.

În fiecare dintre cele două puncte $(\xi_1, \xi_2, \ldots, \xi_{s-1})$ care corespund aceluiaşi punct $(\eta_1, \eta_2, \ldots, \eta_{s-2}, \chi)$, expresia

$$\left|\frac{D(\xi_1, \xi_2, \ldots, \xi_{s-1})}{D(\eta_1, \eta_2, \ldots, \eta_{s-2}, \chi)}\right| = \chi^{s-2}\cdot\frac{\sqrt{p_s\,q_s}}{|D_{s-1}|}$$

are aceeași valoare, deoarece

$$D_{s-1}^2 = q_{s-1}q_s[(p_{s-1}+p_s)(q_{s-1}\eta_{s-1}^2+q_s\eta_s^2)-(\sqrt{p_{s-1}q_{s-1}}\cdot\eta_{s-1}+\sqrt{p_sq_s}\eta_s)^2] =$$

$$= q_{s-1}q_s\left[p_{s-1}+p_s)\left(1-\sum_{h=1}^{s-2}q_h\eta_h^2\right)-\left(\sum_{h=1}^{s-2}\sqrt{p_hq_h}\cdot\eta_h^2\right)\right]$$

și în consecință expresiile $|D_{s-1}|$ sînt unic determinate în funcție de $\eta_1, \eta_2, \ldots, \eta_{s-2}$.

Rezultă că

$$dP = 2\frac{\sqrt{q_1q_2\cdots q_{s-1}}}{(2\pi)^{\frac{s-1}{2}}\sqrt{p_s}}e^{-\frac{\chi^2}{2}}\cdot\chi^{s-2}\cdot\frac{\sqrt{p_sq_s}}{|D_{s-1}|}d\eta_1\cdot d\eta_2\cdot\ldots\cdot d\eta_{s-2}\,d\chi.$$

Deoarece

$$\chi^{s-2}\cdot e^{-\frac{\chi^2}{2}} > 0,$$

pentru orice $\chi > 0$,

$$F(x) = \frac{\int_0^x \chi^{s-2}e^{-\frac{\chi^2}{2}}d\chi}{\int_0^\infty \chi^{s-2}e^{-\frac{\chi^2}{2}}d\chi}$$

este o funcție de distribuție.

Prin schimbarea de variabile $t = \frac{\chi^2}{2}$, găsim

$$\int_0^\infty \chi^{s-2}e^{-\frac{\chi^2}{2}}d\chi = 2^{\frac{s-3}{2}}\int_0^\infty t^{\frac{s-3}{2}}e^{-t}dt = 2^{\frac{s-3}{2}}\Gamma\left(\frac{s-1}{2}\right).$$

Așadar

$$dP = \frac{\chi^{s-2}\cdot e^{-\frac{\chi^2}{2}}\cdot d\chi}{2^{\frac{s-3}{2}}\cdot\Gamma\left(\frac{s-1}{2}\right)}\cdot\frac{\Gamma\left(\frac{s-1}{2}\right)}{\pi^{\frac{s-1}{2}}}\cdot\frac{\sqrt{q_1\cdot q_2\cdots q_s}}{|D_{s-1}|}d\eta_1\cdot d\eta_2\cdots d\eta_{s-2}$$

este produsul a două probabilități elementare independente, dintre care prima este legea lui χ^2.

Introducînd notația

$$\sum_{i=1}^{s}\frac{(\alpha_i-np_i)^2}{np_i} = \chi^2,$$

rezultă că probabilitatea elementară de a avea pentru expresia de mai sus o probabilitate cuprinsă între χ și $\chi + d\chi$ este egală cu

$$\frac{\chi^{s-2}\cdot e^{-\frac{\chi^2}{2}}\cdot d\chi}{2^{\frac{s-3}{2}}\cdot\Gamma\left(\frac{s-1}{2}\right)}. \qquad \text{(VII,6.12)}$$

Se observă ușor că expresia (VII,6.12) este funcția de frecvență $f(x)dx$ a distribuției $H(s-1,1)$ a lui Helmert-Pearson cu $s-1$ grade de libertate ($\sigma = 1$).

Într-adevăr, făcînd substituția $\chi^2 = x$, expresia (VII,6.12) devine

$$f(x)dx = \frac{1}{2^{\frac{s-1}{2}} \cdot \Gamma\left(\frac{s-1}{2}\right)} \cdot x^{\frac{s-1}{2}-1} \cdot e^{-\frac{\chi}{2}} \cdot dx.$$

Notînd valorile empirice prin f_i și cele teoretice cu t_i, $(i = 1, 2,..,s)$, expresia (VII,6.2) se poate scrie

$$\chi^2 = \sum_{i=1}^{s} \frac{(f_i - t_i)^2}{t_i}. \qquad \text{(VII,6.13)}$$

Această expresie urmează tot o lege de probabilitate χ^2 cu $s-1$ grade de libertate; ea constituie, după cum vom vedea, un criteriu foarte important pentru aprecierea semnificației abaterii diferitelor frecvențe statistice, care sînt — după cum știm — de natură empirică, de la frecvențele teoretice corespunzătoare. O metodă simplă pentru aplicarea criteriului χ^2 ne oferă tabelele numerice conținînd diferite valori ale lui χ^2 corespunzătoare probabilităților 0,95, 0,99 etc., pentru diferite numere de grade de libertate.

Mai sus am arătat că pentru o variabilă care ia s valori, numărul gradelor de libertate este egal cu $s-1$.

Dacă avem mai multe variabile parțiale independente, fiecare dintre ele urmînd o lege $H(s_i; \sigma)$, suma lor urmează tot o lege $H(s, \sigma)$, avînd ca număr de grade de libertate suma gradelor de libertate pentru fiecare variabilă parțială.

Această teoremă are o deosebită importanță teoretică și practică.

Dacă două variabile aleatoare independente X_1 și X_2 urmează fiecare respectiv legile $H(s_1, \sigma)$ cu s_1 grade de libertate și $H(s_2, \sigma)$ cu s_2 grade de libertate, atunci suma lor $X_1 + X_2$ urmează legea $H(s_1 + s_2, \sigma)$ cu $s_1 + s_2$ grade de libertate.

Demonstrația rezultă imediat, considerînd funcțiile caracteristice ale celor două variabile X_1 și X_2:

$$\varphi_1(t) = (1 - 2\sigma^2 ti)^{-\frac{s_1}{2}},$$

$$\varphi_2(t) = (1 - 2\sigma^2 ti)^{-\frac{s_2}{2}}.$$

Funcția caracteristică $\varphi(t)$ a variabilei $X_1 + X_2$, în baza teoremei funcției caracteristice a sumei de variabile independente, va fi

$$\varphi(t) = \varphi_1(t) \cdot \varphi_2(t) = (1 - 2\sigma^2 ti)^{-\frac{s_1+s_2}{2}}.$$

Aceasta este expresia funcției caracteristice a legii $H(s_1 + s_2, \sigma)$ cu $s_1 + s_2$ grade de libertate și teorema este astfel demonstrată.

MOMENTELE LEGII $H(s, \sigma)$. 1) *Momente obișnuite.* În teoria probabilităților se demonstrează că primele patru momente obișnuite ale legii $H(s, \sigma)$ au următoarele valori:

$$M_1 = s\sigma^2,$$
$$M_2 = s(s + 2)\sigma^4,$$
$$M_3 = s(s + 2)(s + 4)\sigma^6,$$
$$M_4 = s(s + 2)(s + 4)(s + 6)\sigma^8.$$

2) *Momente centrate.* Avem

$$\mu_1 = 0,$$

$$\mu_2 = M_2 - M_1^2 = 2s\sigma^4.$$

$$\mu_3 = M_3 - 3M_2M_1 + 2M_1^3 = 8s\sigma^6,$$

$$\mu_4 = M_4 - 4M_3M_1 + 6M_2M_1^2 - 3M_1^4 = 12s\sigma^8\,(s+4).$$

Cunoscînd momentele M_1 și μ_2 se poate demonstra că legea $H(s, \sigma)$ se apropie asimptotic de legea normală cînd numărul gradelor de libertate s este foarte mare.

TEOREMĂ. *Dacă* $s \to \infty$, *atunci distribuția* $H(s, \sigma)$ *cu* s *grade de libertate tinde către distribuția normală* $N(0,1)$.

Pentru demonstrarea teoremei ne vom folosi de funcția caracteristică a variabilei reduse

$$\frac{x - M_1(x)}{\sqrt{\mu_2}} = \frac{x - s\sigma^2}{\sqrt{2s\sigma^2}}$$

care, în baza definiției sale, se va scrie

$$\Psi_s(t) = M\left(e^{\frac{x-s\sigma^2}{\sqrt{2s\sigma^2}}ti}\right) = e^{-\sqrt{\frac{s}{2}}ti} \cdot M\left[e^{x\left(\frac{t}{\sqrt{2s\sigma^2}}\right)i}\right] = e^{-\sqrt{\frac{s}{2}}ti} \cdot \varphi\left(\frac{t}{\sqrt{2s\sigma^2}}\right) = e^{-\sqrt{\frac{s}{2}}ti}\left(1 - \sqrt{\frac{s}{2}}ti\right)^{-\frac{s}{2}}.$$

Aplicînd logaritmii în expresia de mai sus, găsim

$$\ln\Psi_s(t) = -\sqrt{\frac{s}{2}}ti - \frac{s}{2}\ln\left(1 - \sqrt{\frac{2}{s}}ti\right).$$

Observăm că logaritmul se poate dezvolta în serie, dacă

$$\left|\sqrt{\frac{2}{s}}t\right| < 1.$$

Deci

$$\ln\Psi_s(t) = -\sqrt{\frac{s}{2}}ti + \frac{s}{2}\left[\sqrt{\frac{2}{s}}ti - \frac{t^2}{2} + O\left(s^{-\frac{3}{2}}\right)\right] = \left[-\frac{t^2}{2} + O\left(s^{-\frac{1}{2}}\right)\right],$$

$O\left(s^{-\frac{1}{2}}\right)$ fiind o funcție asimptotică de ordinul lui $s^{-\frac{1}{2}}$.

La limită, cînd $s \to \infty$, $O\left(s^{-\frac{1}{2}}\right) \to 0$ și, prin urmare, $\Psi_s(t) \to e^{-\frac{t^2}{2}}$ care este funcția caracteristică a legii normale $N(0,1)$.

Se știe că dacă șirul de funcții caracteristice $\Psi_s(t)$ tinde către funcția caracteristică

$$e^{-\frac{t^2}{2}}$$

a legii normale $N(0,1)$, atunci și funcțiile de repartiție corespunzătoare vor tinde, de asemenea către funcția de repartiție a legii normale $N(0,1)$.

Teorema este astfel demonstrată.

7. Legea de probabilitate Student

Dacă considerăm expresia dată de raportul

$$t = \frac{X}{\sqrt{\frac{Y}{s}}}, \qquad \text{(VII,7.1)}$$

în care X este o variabilă independentă avînd legea de probabilitate de tipul $N(0, \sigma)$, iar Y o altă variabilă independentă, avînd legea de probabilitate $H(s, \sigma)$ legea de probabilitate a acesteia este de forma

$$Y = Y_0 \cdot \left(1 + \frac{t^2}{s}\right)^{-\frac{s+1}{2}}$$

sau

$$Y = \frac{Y_0}{\left(1 + \frac{t^2}{s}\right)^{\frac{s+1}{2}}}. \qquad \text{(VII,7.2)}$$

Probabilitatea elementară pentru punctul (X, Y) este

$$\frac{1}{\sqrt{\pi} \cdot 2^{\frac{s+1}{2}} \sigma^{s+1} \cdot \Gamma\left(\frac{s}{2}\right)} Y^{\frac{s}{2}-1} \cdot e^{-\frac{X^2+Y}{2\sigma^2}} \cdot dX \cdot dY,$$

Efectuînd schimbarea de variabile

$$X = t\sqrt{\frac{u}{s}} \text{ şi } Y = u,$$

determinantul funcţional al transformării are valoarea

$$\frac{D(X, Y)}{D(t, u)} = \begin{vmatrix} \sqrt{\frac{u}{s}} & \frac{t}{2\sqrt{us}} \\ 0 & 1 \end{vmatrix} = \sqrt{\frac{u}{s}}.$$

Probabilitatea elementară pentru punctul (X, Y) devine

$$\frac{1}{\sqrt{\pi s} \cdot 2^{\frac{s+1}{2}} \cdot \sigma^{s+1} \cdot \Gamma\left(\frac{s}{2}\right)} u^{\frac{s-1}{2}} \cdot e^{-u\frac{\frac{t^2}{s}+1}{2\sigma^2}} \cdot dt \cdot du.$$

Pentru a găsi distribuţia marginală a lui t, vom integra expresia de mai sus în raport cu variabila u între limitele 0 şi ∞. Astfel vom avea

$$g(t)dt = \frac{dt}{\sqrt{\pi s} \cdot 2^{\frac{s+1}{2}} \cdot \sigma^{s+1} \cdot \Gamma\left(\frac{s}{2}\right)} \int_0^\infty u^{\frac{s-1}{2}} e^{-u\frac{\frac{t^2}{s}+1}{2\sigma^2}} du.$$

Ținînd seama de definiția funcției $\Gamma(x)$ deducem

$$\int_0^\infty u^{\frac{s-1}{2}} \cdot e^{-u\frac{1+\frac{t^2}{s}}{2\sigma^2}} \cdot du = \frac{\Gamma\left(\frac{s+1}{2}\right)}{\left(\frac{1+\frac{t^2}{s}}{2\sigma^2}\right)^{\frac{s+1}{2}}} = 2^{\frac{s+1}{2}} \cdot \sigma^{s+1} \cdot \Gamma\left(\frac{s+1}{2}\right)\left(1+\frac{t^2}{s}\right)^{-\frac{s+1}{2}}.$$

Prin urmare

$$g(t) = \frac{\Gamma\left(\frac{s+1}{2}\right)}{\sqrt{\pi s}\cdot\Gamma\left(\frac{s}{2}\right)} \cdot \left(1+\frac{t^2}{s}\right)^{-\frac{s+1}{2}},$$

care este funcția de frecvență a distribuției variabilei t.

Legea de probabilitate dată de expresia (VII,7.2) este cunoscută sub numele de *legea Student* *) *cu s grade de libertate sau legea t cu s grade de libertate.*

Legea Student se aplică în cazul selecțiilor cu volum mic. În teoria selecției, această lege este cunoscută sub forma

$$t = \frac{\widetilde{X} - M}{\sigma' : \sqrt{N-1}} \qquad \text{(VII,7.3)}$$

în care: $\widetilde{X} - M$ reprezintă abaterea mediei de selecție de la media colectivității generale, avînd legea de probabilitate normală de tipul $N(0,1)$; σ' abaterea standard, a selecției, iar N volumul selecției; variabila de la numitor $\frac{\sigma'}{\sqrt{N-1}}$ are legea de probabilitate χ^2.

O proprietate remarcabilă a legii de probabilitate a variabilei t, (relația (VII,7.2)), constă în faptul că aceasta *depinde numai de parametrii selecției*, calculați nemijlocit pe baza datelor observării selective, fiind independentă de dispersia din colectivitatea generală.

Din graficul din figura 10 se poate observa că *în cazul cînd numărul gradelor de libertate s crește, legea de probabilitate a variabilei t se apropie de legea normală.*

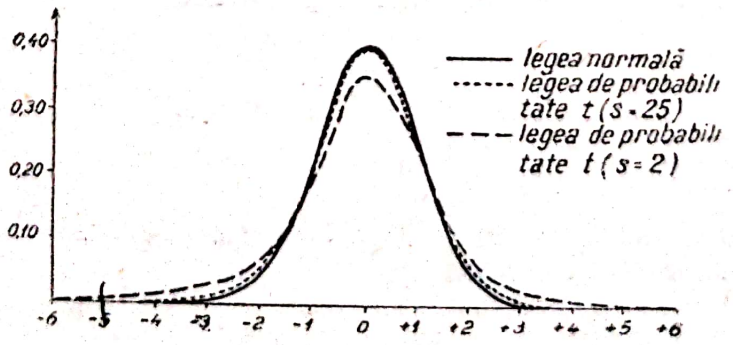


Fig. 10

*) Numele de „Student" este cunoscutul pseudonim al matematicianului V. S. Gosset (Anglia).

Momente

O particularitate interesantă a legii Student constă în aceea că momentele ei obişnuite coincid cu momentele centrate, deoarece momentul de ordinul întîi este egal cu zero.

Pentru momentele centrate de ordin par ale legii Student, rezultă următoarele valori :

$$\mu_2 = \frac{s}{s-2},$$

$$\mu_4 = \frac{3s^2}{(s-2)(s-4)},$$

$$\mu_6 = \frac{15s^3}{(s-2)(s-4)(s-6)}.$$

. . . . . .

Am arătat că funcţia de frecvenţă a distribuţiei Student cu s grade de libertate este

$$g(t) = \frac{\Gamma\left(\frac{s+1}{2}\right)}{\sqrt{\pi s}\,\Gamma\left(\frac{s}{2}\right)}\left(1+\frac{t^2}{s}\right)^{-\frac{s+1}{2}}.$$

În cazul acestei distribuţii observăm că momentele obişnuite coincid cu momentele centrate, deoarece momentul de ordinul întîi este egal cu zero.

Avem evident,

$$m_{2r+1} = \frac{\Gamma\left(\frac{s+1}{2}\right)}{\sqrt{\pi s}\cdot\Gamma\left(\frac{s}{2}\right)}\int_{-\infty}^{+\infty} t^{2r+1}\left(1+\frac{t^2}{s}\right)^{-\frac{s+1}{2}} dt = 0,$$

deoarece funcţia $t^{2r+1}\left(1+\frac{t^2}{s}\right)$ este impară şi

$$m_{2r} = \frac{\Gamma\left(\frac{s+1}{2}\right)}{\sqrt{\pi s}\cdot\Gamma\left(\frac{s}{2}\right)}\int_{-\infty}^{+\infty} t^{2r}\left(1+\frac{t^2}{s}\right)^{-\frac{s+1}{2}} dt.$$

Folosind substituţia $\frac{t^2}{s} = x$, momentul centrat de ordinul $2r$ este

$$m_{2r} = \frac{s^r\Gamma\left(\frac{s+1}{2}\right)}{2\sqrt{\pi}\cdot\Gamma\left(\frac{s}{2}\right)}\int_{-\infty}^{+\infty} x^{r-\frac{1}{2}}(1+x)^{-\frac{s+1}{2}} dx.$$

Se ştie însă că

$$\int_0^{\infty} x^{p-1}(1+x)^{-(p+q)} dx = \frac{\Gamma(p)\Gamma(q)}{\Gamma(p+q)}.$$

Atunci rezultă că

$$\int_{-\infty}^{+\infty} x^{r-\frac{1}{2}}(1+x)^{-\frac{s+1}{2}} dx = 2\frac{\Gamma\left(r+\frac{1}{2}\right)\cdot\Gamma\left(\frac{s}{2}-r\right)}{\Gamma\left(\frac{s+1}{2}\right)}.$$

Prin urmare, ținînd seama de relațiile găsite

$$m_{2r} = \frac{s^r\,\Gamma\left(\frac{s+1}{2}\right)}{2\sqrt{\pi}\,\Gamma\left(\frac{s}{2}\right)} \cdot \frac{2\,\Gamma\left(r+\frac{1}{2}\right)\Gamma\left(\frac{s}{2}-r\right)}{\Gamma\left(\frac{s+1}{2}\right)} = \frac{s^r}{\sqrt{\pi}} \cdot \frac{\Gamma\left(r+\frac{1}{2}\right)\Gamma\left(\frac{s}{2}-r\right)}{\Gamma\left(\frac{s}{2}\right)}.$$

Observăm că momentul centrat m_{2r} există numai dacă $r < \frac{s}{2}$.

Expresia momentului centrat de ordinul $2r$ se poate pune sub o formă mai simplă. Într-adevăr, folosind o proprietate cunoscută a funcției $\Gamma(x)$, avem

$$\Gamma\left(\frac{s}{2}\right) = \left(\frac{s}{2}-1\right)\left(\frac{s}{2}-2\right)\cdots\left(\frac{s}{2}-r\right)\Gamma\left(\frac{s}{2}-r\right).$$

La fel, după aceeași formulă,

$$\Gamma\left(r+\frac{1}{2}\right) = \left(r-\frac{1}{2}\right)\left(r-\frac{3}{2}\right)\cdots\frac{1}{2}\cdot\Gamma\left(\frac{1}{2}\right).$$

Înlocuind cu formula găsită valorile de mai sus ale lui $\Gamma\left(\frac{s}{2}\right)$ și $\Gamma\left(r+\frac{1}{2}\right)$, obținem

$$m_{2r} = \frac{s^r}{\sqrt{\pi}} \cdot \frac{\left(r-\frac{1}{2}\right)\left(r-\frac{3}{2}\right)\cdots\frac{1}{2}\sqrt{\pi}\,\Gamma\left(\frac{s}{2}-r\right)}{\left(\frac{s}{2}-1\right)\left(\frac{s}{2}-2\right)\cdots\left(\frac{s}{2}-r\right)\Gamma\left(\frac{s}{2}-r\right)} = \frac{s^r\cdot 1\cdot 3\cdots(2r-3)(2r-1)}{(s-2)(s-4)\cdots(s-2r)}.$$

Particularizînd pe r, deducem momentele centrate de ordin par ale distribuției Student

$$\mu_2 = \frac{s}{s-2},$$

$$\mu_4 = \frac{3s^2}{(s-2)(s-4)},$$

$$\mu_6 = \frac{15s^3}{(s-2)(s-4)(s-6)}.$$

Putem acum demonstra riguros și proprietatea de convergență pe care am semnalat-o.

TEOREMĂ. *Dacă s, numărul gradelor de libertate, crește nemărginit, atunci distribuția Student cu s grade de libertate tinde către distribuția normală $N(0,1)$.*

Pentru demonstrarea teoremei este suficient să arătăm că șirul de momente m_p, oricare ar fi p întreg și pozitiv, tinde către momentele distribuției normale $N(0,1)$, dacă $s \to \infty$.

În acest caz, funcțiile de repartiție corespunzătoare tind, de asemenea, către funcția de repartiție a distribuției normale $N(0,1)$.

Dacă p este impar, atunci momentele distribuției Student sînt egale cu momentele distribuției normale.

Dacă p este, însă, par, obținem

$$\lim_{s\to\infty} m_{2r} = \lim_{s\to\infty} \frac{s^r\cdot 1\cdot 3\cdots(2r-3)(2r-1)}{(s-2)(s-4)\cdots(s-2r)} = 1\cdot 3\cdot 5\cdots(2r-1),$$

întrucît numărătorul și numitorul sînt polinoame în s de gradul r cu coeficienții respectiv egali cu $1\cdot 3\cdot 5\cdots(2r-1)$ și 1.

Cum, însă, $1\cdot 3\cdot 5\cdots(2r-1)$ este momentul de ordinul $2r$ a distribuției normale $N(0,1)$, teorema este demonstrată.

Scanned with OKEN Scanner

Capitolul VIII

Legea numerelor mari

1. Considerații generale

Încă la capitolul I, cînd am definit noțiunea de probabilitate, am precizat că probabilitatea nu este o valoare teoretică ruptă de realitate, fictivă, obținută prin calcule formale, ci că probabilitatea este legată de frecvența statistică a fenomenelor de masă, deoarece s-a constatat că la repetarea experiențelor în sistemul de condiții A, frecvențele se grupează în jurul probabilității p, care este o constantă obiectivă, depinzînd de natura fenomenului considerat și de sistemul de condiții A.

Adîncind acum aceste considerații putem releva faptul că multe fenomene din natură și din viața social-economică se bucură de o anumită *regularitate*, de așa-numita „stabilitate statistică" care se manifestă cu atît mai pronunțat, cu cît numărul elementelor componente ale fenomenului studiat este mai mare. Pot fi date numeroase exemple de asemenea regularități care au atras atenția oamenilor de știință, găsindu-și apoi și aplicații practice, înainte chiar de a fi fost studiate din punct de vedere matematic.

S-a remarcat, de pildă, că raportul dintre numărul copiilor de sex masculin nou-născuți și al acelora de sex feminin se păstrează în fiecare an, aproximativ același. De asemenea, în experiența clasică a aruncării unei monede, raportul dintre numărul cazurilor în care se prezintă o anumită față și numărul total al aruncărilor, este cu regularitate egal, aproximativ, cu $\frac{1}{2}$, care reprezintă o valoare teoretică constantă, obiectivă, a probabilității p de a se verifica apariția uneia din fețe.

Tocmai această regularitate, stabilitate statistică sau *legitate statistică* cum i se mai spune *), am pus-o la baza definiției probabilității, ară-

*) Vezi *Bazele filozofiei marxiste*. Ed. Politică, București, 1959, Cap. VI, p. 252.

tînd că fenomenele la care se aplică calculul probabilităților trebuie să fie *stabile*, în sensul pe care l-am arătat.

Am văzut că și axiomele teoriei probabilităților date de Kolmogorov, întocmai ca și axiomele celorlalte discipline matematice, nu sînt arbitrare, ci rezultă din studiul îndelungat al diferitelor fenomene și din cunoașterea legilor lor de dezvoltare.

Prin urmare, la definirea probabilității — și acest lucru l-am mai subliniat cu alt prilej — în accepția ei modernă, s-a ajuns pe calea statistică.

Faptul că între frecvență și probabilitate există anumite legături ce se pot exprima matematic, aceasta nu înseamnă, însă, că stabilitatea statistică a fenomenelor poate fi demonstrată matematic.

Pe de altă parte, nici concepția lui Mises, după care relația

$$\lim_{n\to\infty} f_n = p \qquad \text{(VIII,1.1)}$$

este dată drept definiție a probabilității, nu este întemeiată din punct de vedere științific.

O regularitate statistică nu se poate demonstra după cum, din punct de vedere matematic, nu se poate demonstra, nici regularitatea temperaturii de fierbere a apei la 100°C. Este vorba, deci, de un adevăr constatat experimental, un postulat, care este folosit atît în aplicații practice cît și în cercetări științifice, teoretice.

Tocmai de aceea, studiul aprofundat al frecvențelor statistice, al convergenței șirului de frecvențe prezintă un interes deosebit, constituind în prezent unul dintre cele mai importante capitole ale calculului probabilităților.

În teoria probabilităților se întîlnesc diferite moduri de convergență, exprimînd, la limită, legătura strînsă între frecvență și probabilitate.

1°. Să considerăm șirul de frecvențe relative,

$$f_1,\ f_2,\dots,\ f_n^{\cdot}.$$

Dacă există o constantă p, astfel încît să avem

$$\lim_{n\to\infty} f_n^{\cdot} = p$$

spunem că șirul tinde către p în mod *sigur* sau în *sensul analizei matematice clasice*.

Sîntem în prezența celei mai strînse legături care poate exista între frecvența $f_n^{\cdot}$ și probabilitatea p.

2°. Între frecvența f_n și probabilitatea p există însă, uneori, o relație de natură cu totul deosebită, statistică. Această relație este specifică teoriei probabilităților.

Pe baza ei afirmăm că probabilitatea inegalității

$$|f_n - p| < \varepsilon \qquad \text{(VIII,1.2)}$$

în care ε este un număr dat, pozitiv și arbitrar de mic, tinde către 1 cînd n crește necontenit.

Scanned with OKEN Scanner

Cu alte cuvinte, avem proprietatea de limită

$$\lim_{n\to\infty} P(|f_n - p| < \varepsilon) = 1. \qquad \text{(VIII,1.3)}$$

Limita de mai sus are loc *în probabilitate*. Şirul frecvenţelor tinde către p în sensul calculului probabilităţilor.

3°. O altă convergenţă, considerată pentru prima oară de Cantelli, este *convergenţa puternică*.

Se spune că şirul f_n admite o convergenţă puternică către p, dacă probabilitatea simultană a inegalităţilor

$$|f_n - p| < \varepsilon, \ldots |f_{n+k} - p| < \varepsilon \qquad \text{(VIII,1.4)}$$

tinde către 1, cînd n creşte necontenit; în această inegalitate ε este un număr pozitiv dat, arbitrar, iar k un număr întreg pozitiv.

2. Teorema lui Bernoulli

J. Bernoulli a formulat o teoremă interesantă, legată de urna care-i poartă numele. Prin conţinutul ei, această teoremă exprimă una dintre cele mai importante legi ale teoriei probabilităţilor, cunoscută sub numele de *legea numerelor mari*. Mai precis, ea exprimă unul din aspectele acestei legi. Iată cum se poate enunţa teorema lui Bernoulli, concepută încă la începutul secolului al XVIII-lea:

Intr-o serie de probe, repetate de un număr mare de ori n, frecvenţa relativă f_n a evenimentului A, a cărui probabilitate de verificare este p, tinde către p „în probabilitate".

Dacă notăm cu α numărul care arată de cîte ori s-a realizat A în n probe consecutive, inegalitatea

$$|f_n - p| < \varepsilon$$

în care ε este un număr pozitiv, arbitrar de mic, este îndeplinită în acelaşi timp cu inegalitatea

$$(\alpha - np) < n\varepsilon \qquad \text{(VIII,2.1)}$$

deoarece

$$f_n = \frac{\alpha}{n}.$$

Dacă notăm cu P_n probabilitatea inegalităţii (VIII, 2.1) şi aplicăm teorema lui Cebîşev, avem

$$P_n \geqslant 1 - \frac{\sigma^2}{n^2\varepsilon^2},$$

în care σ^2 reprezintă dispersia variabilei considerate și cum în cazul urnei Bernoulli dispersia este egală cu npq, putem scrie

$$P_n \geqslant 1 - \frac{npq}{n^2\varepsilon^2} = 1 - \frac{pq}{n\varepsilon^2}.$$

Rezultă că, dacă n crește necontenit, P_n tinde către 1. Or, tocmai acest lucru se afirmă și în teorema lui Bernoulli.

3. Teorema lui Poisson

Considerăm șirul de evenimente

$$E_1, E_2, \ldots, E_n, \ldots$$

a căror probabilități de verificare au valorile succesive

$$p_1, p_2, \ldots, p_n, \ldots$$

Notînd cu f_n frecvența relativă a numărului care indică de cîte ori s-au realizat evenimentele $E_1, E_2, \ldots, E_n, \ldots$ și cu P expresia

$$P = \lim_{n\to\infty} \frac{P_1 + P_2 + \ldots + P_n}{n}, \qquad \text{(VIII,3.1)}$$

Poisson a demonstrat — presupunînd că limita de mai sus există — că f_n tinde către P *în probabilitate*.

Teorema lui Poisson generalizează teorema lui Bernoulli.

Demonstrarea teoremei lui Poisson se face pe aceeași cale ca și pentru teorema lui Bernoulli. De aceea o lăsăm pe seama cititorului.

4. Teorema lui Cantelli

Cantelli a arătat că variabilele care schematic se pot reprezenta printr-o urnă a lui Bernoulli admit și o *convergență puternică* în sensul definiției pe care am dat-o.

Această proprietate fundamentează aplicarea practică a teoriei probabilistice referitoare la urna lui Bernoulli.

Într-adevăr, pentru scopuri practice, o probabilitate foarte apropiată de 1 poate fi considerată drept certitudine.

De exemplu, dacă într-o urnă avem un milion de bile albe și una neagră, probabilitatea să scoatem o bilă albă este

$$\frac{10^6}{10^6+1} = 1 - \frac{1}{10^6+1}.$$

Această probabilitate este foarte aproape de unu însă, din punct de vedere practic, o putem considera chiar egală cu 1, în sensul că

exprimă probabilitatea unui eveniment de realizare sigură din punct de vedere experimental.

Aşadar, dacă în baza teoremei lui Bernoulli, inegalitatea

$$| f_n - p | < \varepsilon$$

se prezintă cu o probabilitate atît de apropiată de 1 cît vrem noi, cu condiţia ca $n \geqslant N$, rezultă că, din punct de vedere practic, putem să o considerăm ca o certitudine.

În diferite aplicaţii ale acestor şiruri de variabile este însă necesar să se verifice nu numai inegalitatea de mai sus, ci toate inegalităţile care o urmează :

$$| f_n - p | < \varepsilon, \ldots, \quad | f_{n+k} - p | < \varepsilon, \ldots \qquad (H)$$

Prin urmare, trebuie să existe o convergenţă puternică către p a şirului de frecvenţe f_n.

Vom expune pe scurt demonstraţia acestei proprietăţi dată de Cantelli.

Dacă notăm cu $1-q_i$ probabilitatea inegalităţii $| f_i - p | < \varepsilon$, ştim că în baza teoremei lui Boole

$$p_n \geqslant 1 - (q_n + q_{n+1} + \ldots),$$

în cazul cînd notăm cu p_n probabilitatea realizării evenimentului (H).

Este suficient ca seria

$$q_1 + q_2 + \ldots + q_n + \ldots$$

să fie convergentă, pentru ca p_n să tindă către 1, cînd n creşte nemărginit.

În baza unei formule a lui Medolaghi, care generalizează inegalitatea lui Cebîşev, am văzut că

$$q_i \leqslant \frac{M_4^{(i)}}{\varepsilon^4},$$

unde $M_4^{(i)}$ este momentul de ordinul al patrulea al abaterii relative $f_i - p$.

Însă

$$M_4^{(i)} = M(f_i - p)^4 = \frac{1}{i^2}\left(3p^2q^2 - \frac{6p^2q^2}{i} + \frac{pq}{i}\right)$$

şi

$$3p^2q^2 - \frac{6p^2q^2}{i} + \frac{pq}{i} \leqslant 3p^2q^2 + \frac{pq}{i} \leqslant \frac{3}{16} + \frac{1}{4i} < \frac{3}{16} + \frac{1}{4} = \frac{7}{16} < 1$$

deoarece

$$pq \leqslant \frac{1}{4}.$$

Prin urmare,

$$M_4^{(i)} = \frac{1}{i^2}\left[3p^2q^2 - \frac{6p^2q^2}{i} + \frac{pq}{i}\right] \leqslant \frac{1}{i^2}.$$

În baza inegalității lui Medolaghi, rezultă că

$$q_i \leqslant \frac{1}{\varepsilon^4 i^2}$$

Această relație ne arată, însă, că seria q_n este convergentă și, prin urmare, teorema este demonstrată.

A. I. Hincin a dat acestei proprietăți de convergență o interesantă generalizare.

El a considerat un șir de numere pozitive

$$\varepsilon_1, \ldots, \varepsilon_n, \ldots$$

tinzînd către zero, cînd n tinde către infinit și și-a propus să le determine astfel încît probabilitatea inegalităților

$$|f_n - p| < c\varepsilon_n, \ldots, |f_{n+k} - p| < c\varepsilon_{n+k}$$

să tindă către 1, pentru $c < 1$, cînd n crește nemărginit și să tindă către zero, în aceleași condiții, dacă $c > 1$.

Rezultatul important stabilit de Hincin este că ε_n are valoarea

$$\varepsilon_n = \sqrt{n \lg \lg n} \cdot$$

5. Un exemplu de convergență sigură *)

Sînt unele fenomene statistice a căror frecvență, la limită, în sensul analizei este egală cu probabilitatea. În asemenea cazuri avem stabilitate matematică a frecvenței acestor fenomene.

Să considerăm o urnă care conține m bile albe și m bile negre; din această urnă se execută o serie de extracții, înlocuind de fiecare dată bila extrasă printr-o bilă de cealaltă culoare. În aceste condiții, probabilitatea ca în extracția de rangul n să scoatem o bilă albă este, din motive de simetrie, egală cu probabilitatea de a scoate o bilă neagră. Prin urmare această probabilitate este egală cu $\frac{1}{2}$.

Ne propunem să arătăm că

$$\lim_{n \to \infty} f_n = \frac{1}{2},$$

prin f_n înțelegînd frecvența relativă a numărului de bile albe ieșite în n extracții.

Să notăm prin α numărul de bile albe ieșite în n extracții. Frecvența relativă are valoarea

$$f_n = \frac{\alpha}{n} \cdot$$

*) O. Onicescu și Gh. Mihoc, *Sopra le leggi-limite delle probabilita.* Giornale dell'Istituto italiano degli Attuari, VII, *1* (1936).

Să stabilim acum numărul de bile albe rămase în urnă după efectuarea celor n extracţii, în condiţiile precizate mai sus. Acest număr este, evident, egal cu

$$(m-\alpha)+(n-\alpha)=m+n-2\alpha.$$

Dat fiind că numărul total al bilelor din urnă rămîne întotdeauna acelaşi şi egal cu $2m$, este clar că numărul bilelor albe rămase în urnă după n extracţii în care au apărut α bile albe va fi cuprins între 0 bile albe şi $2m$ bile albe, adică

$$0 \leqslant m+n-2\alpha \leqslant 2m$$

sau

$$-m \leqslant n-2\alpha \leqslant m$$

Împărţind prin $2n$, obţinem

$$-\frac{m}{2n} \leqslant \frac{1}{2}-\frac{\alpha}{n} \leqslant \frac{m}{2n}$$

sau

$$-\frac{m}{2n} \leqslant \frac{1}{2}-f_n \leqslant \frac{m}{2n}.$$

Ultimele inegalităţi ne arată că, dacă n tinde către infinit,

$$\lim_{n\to\infty} f_n = \frac{1}{2}.$$

A rezultat, deci, că frecvenţa relativă f_n, deşi oscilează, tinde totuşi către probabilitatea $\frac{1}{2}$, în sensul clasic al analizei matematice.

Se poate considera şi o schemă mai generală care ne duce la acelaşi rezultat. O asemenea schemă a fost arătată la capitolul I, odată cu prezentarea metodei funcţiilor arbitrare a lui Hincin.

6. Convergenţa variabilei aleatoare în general

Convergenţele în probabilitate şi în sens ordinar exprimate prin teoremele de la punctele precedente nu sînt decît cazuri particulare de convergenţă, deoarece în cazul general un şir convergent de variabile nu tinde către o constantă, ci tot către o variabilă.

Considerînd, de exemplu, şirul

$$X_1,\ X_2,\ldots,\ X_n,\ldots$$

spunem prin definiţie că acest şir tinde *în probabilitate* către variabila X, dacă probabilitatea inegalităţii

$$|X_n - X| < \varepsilon,$$

unde ε este un număr pozitiv, arbitrar de mic, tinde către 1, cînd n crește necontenit.

Spunem că două șiruri de variabile aleatoare

$$X_1,\ X_2,..,\ X_n,...$$
$$Y_1,\ Y_2,...,\ Y_n,...$$

sînt *echivalente*, dacă

$$\lim_{n\to\infty} (X_n - Y_n) = 0.$$

În mod analog se definește echivalența puternică și echivalența în probabilitate. În legătură cu aceste moduri de convergență se pot pune diverse probleme noi și interesante asupra cărora, însă, nu ne vom opri aici.

7. Esența legii numerelor mari. Cîteva observații

Expunerea de mai sus a proprietăților de convergență ne permite să reținem, în concluzie, că *toate aceste proprietăți formează împreună legea numerelor mari* și că fiecare în parte relevă un aspect deosebit, interesant, al acestei legi.

Cu excepția cazului general al proprietății de convergență a unei variabile care tinde tot către o variabilă, celelalte proprietăți referitoare la legea numerelor mari ne-au învederat existența unor șiruri de variabile care tind către o constantă în probabilitate. Tocmai în această tendință se manifestă acțiunea obiectivă a legii numerelor mari, a creșterii necontenite, a numărului cazurilor observate. În aceasta constă esența acestei legi.

Legea numerelor mari este una din *legitățile statistice* a căror esență constă în aceea că sînt valabile pentru un *întreg ansamblu de fenomene* și nu pentru fiecare fenomen separat al acestui ansamblu. Ea are toate caracterele legii, obiectivitate, necesitate, generalitate și se extinde asupra tuturor ansamblurilor de același tip. Exprimînd legătura esențială și necesară a fenomenelor, pentru descoperirea ei trebuie examinat un număr foarte mare de variații întîmplătoare ale fenomenelor studiate.

În concepția filozofiei marxiste, legea numerelor mari exprimă dialectica dintre necesitate și întîmplare; în virtutea acestei legi, acțiunea de ansamblu a unui mare număr de factori întîmplători duce la un rezultat care nu depinde de un caz particular *).

Pe marginea proprietăților de limită referitoare la legea numerelor mari se impun cîteva observații și precizări:

1°. Dacă ne referim, de pildă, la teorema lui Bernoulli, o judecată superficială ar putea duce la concluzia că teorema lui Bernoulli demonstrează legea statistică a întîmplării. O asemenea concluzie ar fi cu totul greșită.

Teorema lui Bernoulli nu demonstrează că inegalitatea

$$|f_n - p| < \varepsilon$$

*) Vezi în acest sens *Bazele filozofiei marxiste*. Ed. politică, București, 1960, Cap. VI, p. 253.

este îndeplinită pentru n mare ci numai că are şansă să fie realizată experimental şi aceasta, în baza faptului că admite o probabilitate foarte apropiată de 1.

Am mai precizat şi la începutul acestui capitol, că realizarea aproape sigură a unui eveniment a cărui probabilitate este foarte apropiată de 1 nu se poate demonstra teoretic, ea constituind pentru o anumită categorie de fenomene o lege obiectivă care se verifică experimental după cum se verifică întotdeauna faptul că apa fierbe la 100°C sau că un metal, în anumite condiţii, se topeşte la N°C.

Prin urmare, teorema lui Bernoulli nu demonstrează stabilitatea fenomenelor.

2°. Nu trebuie să tragem concluzia că legea numerelor mari exprimată printr-o convergenţă în sensul analizei clasice — convergenţă ordinară — sau printr-o convergenţă în probabilitate, constituie o proprietate caracteristică a oricăror şiruri de variabile statistice. O asemenea concluzie ar fi şi ea total greşită. Se pot da exemple în care legea numerelor mari nu mai acţionează, întrucît încetează de a mai exista sub orice formă.

O schemă simplă, datorită lui Markov *), lămureşte această problemă:

Se porneşte de la o urnă conţinînd două bile, una albă şi una neagră şi se consideră ipoteza că, după fiecare extracţie, bila ieşită este înlocuită cu două bile de aceeaşi culoare. După n extracţii putem scoate

$$0,\ 1,\ 2,\ \ldots,\ n$$

bile albe, cu probabilităţile respective

$$P_{n(0)},\ P_{n(1)},\ P_{n(2)},\ \ldots,\ P_{n(n)}:$$

Din ipoteza considerată, rezultă relaţia de recurenţă

$$P_{n(\alpha)} = \frac{\alpha}{n+1}\, P_{n-1(\alpha-1)} + \frac{n-\alpha}{n+1}\, P_{n-1(\alpha-1)}.$$

Presupunînd că

$$P_{n-1(\alpha-1)} = P_{n-1(\alpha-1)} = \frac{1}{n},$$

din egalitatea precedentă obţinem

$$P_{n(\alpha)} = \frac{1}{n+1};$$

Prin urmare, dacă presupunem că

$$P_{n-1(\alpha-1)} = \frac{1}{n}$$

oricare ar fi α, rezultă că

$$P_{n(\alpha)} = \frac{1}{n+1}, \qquad (\alpha = 0,1,2,\ \ldots,\ n).$$

*) A. A. Markov (1856—1922), ilustru matematician rus, a lucrat în domeniul analizei matematice, teoriei numerelor şi teoriei probabilităţilor. A. A. Markov a definit şi a studiat pentru prima oară şirurile de variabile dependente care îi poartă numele şi stau astăzi la baza teoriei moderne a proceselor stohastice. Tot Markov a fundamentat probabilistic metoda celor mai mici pătrate.

Scanned with OKEN Scanner

Știind că

$$P_{n(0)} = P_{1(1)} = \frac{1}{2},$$

rezultă că relația

$$P_{n(\alpha)} = \frac{1}{n+1}$$

este adevărată, oricare ar fi n și α, cuprins între 0 și n inclusiv.

Probabilitățile ca din n extracții să se obțină succesiv frecvențele

$$\frac{0}{n}, \frac{1}{n}, \frac{2}{n}, \ldots, \frac{n}{n}$$

sînt respectiv

$$\frac{1}{n+1}, \frac{1}{n+1}, \ldots, \frac{1}{n+1}.$$

Nici una dintre frecvențele de mai sus nu se prezintă cu o probabilitate mai mare; toate sînt egal probabile și, prin urmare nu poate fi vorba de vreo convergență, de orice natură.

Exemplul lui Markov este deosebit de instructiv pentru statisticieni. El arată că în acest caz nu putem face nici un fel de previziune asupra frecvențelor relative după un număr mare de experiențe. Oricare frecvență are aceeași șansă să se prezinte ca și celelalte. Este inutil să facem o statistică a modurilor experimentale de realizare ale acestei experiențe. Nu vom putea trage nici o concluzie utilă pentru studiul fenomenului considerat.

Calculul probabilităților oferă concepția matematică de bază a statisticii matematice. În acest scop, legea numerelor mari ne arată modul și măsura în care putem trage concluzii statistice valabile referitoare la anumite categorii de fenomene.

În celelalte capitole, în care vom trata statistica ca disciplină științifică de sine stătătoare, vom vedea cum se realizează acest lucru.

Scanned with OKEN Scanner

PARTEA A DOUA

ELEMENTE DE ANALIZĂ A DISTRIBUŢIILOR STATISTICE

Scanned with OKEN Scanner

Scanned with OKEN Scanner

Capitolul IX

Distribuţii statistice cu o singură caracteristică

1. Noţiuni introductive. Caracteristici variabile

Din cursurile de statistică teoretică se ştie că după epuizarea primei etape a oricărei cercetări statistice — observarea statistică — urmează procesul complex al centralizării şi grupării datelor primare culese, folosindu-se cu acest prilej procedee specifice statisticii, diferenţiate în funcţie de esenţa fenomenelor supuse cercetării. Centralizarea şi gruparea statistică se fac după un plan fundamentat ştiinţific conceput iniţial care, împreună cu programul şi planul organizatoric al observării statistice, constituie o unitate. Sînt cunoscute o serie de principii care trebuie respectate cu stricteţe pentru a se putea obţine o sistematizare şi generalizare ştiinţifică a materialului statistic, cules în etapa observării. Aşa, de exemplu, orice statistician ştie că la baza centralizării ştiinţifice stă *gruparea* fenomenelor în studiu, pe bază de caracteristici esenţiale şi că în statistica socialistă gruparea reprezintă metoda fundamentală a oricărei cercetări statistice. Privită dintr-un anumit punct de vedere, strict tehnic, gruparea înseamnă împărţirea materialului statistic în anumite grupe omogene, după un plan stabilit în prealabil, întemeiat pe o caracteristică oarecare de grupare.

Problema grupării nu este, însă, o problemă de strictă specialitate îngust tehnică; o grupare ştiinţifică — aşa cum indica Lenin — are rolul de a face posibilă obţinerea unei caracterizări multilaterale a întregii colectivităţi studiate, după *tipurile ei calitative*. Cu alte cuvinte, gruparea are rolul de a diferenţia *grupe calitativ omogene* în cadrul colectivităţii totale *) care constituie obiectul cercetării. Astfel înţeleasă, gruparea este nu numai baza ştiinţifică a prelucrării materialului statistic al observării, ci şi una

*) În locul acestei noţiuni vom folosi şi noţiunea de „*colectivitate generală*", îndeosebi atunci cînd vor fi expuse problemele esenţiale ale metodei selective, opunînd-o noţiunii de „*colectivitate de selecţie*".

din metodele proprii ale statisticii — în sensul ei clasic de știință socială — de cunoaștere științifică a fenomenelor și proceselor social-economice.

Mai trebuie precizat, de altfel, că metoda grupării nu se poate lega de un anumit stadiu al cercetării statistice, deși, în scop didactic, ea este prezentată de obicei într-un capitol separat din cadrul expunerii asupra prelucrării statistice; la fel ca și analiza politico-economică, gruparea se aplică în toate etapele cercetării statistice.

În lucrările sale practice și teoretice, statistica socialistă acordă o importanță deosebită metodei grupărilor ca instrument de analiză și de cunoaștere, folosind în acest scop indicațiile teoretice și practice ale clasicilor marxism-leninismului. Sarcina principală pe care o rezolvă metoda grupării — aceea de a separa și caracteriza *tipurile* și *formele* social-economice ale fenomenelor și proceselor studiate — se bazează deci, nemijlocit, pe cunoașterea și aplicarea justă a tezelor teoriei marxist-leniniste, potrivit cărora pentru înțelegerea științifică a fenomenelor sociale este necesar să pornim de la analiza relațiilor de producție și, în primul rînd, de la analiza formelor de proprietate asupra mijloacelor de producție.

În cele ce urmează, nu ne vom opri, desigur, asupra problemelor multiple pe care le ridică procesul de separare a grupelor omogene din punct de vedere social-economic și, în general, asupra problemelor cu caracter social-economic ce se ridică în toate etapele cercetării statistice, preocuparea noastră principală fiind aceea de a scoate în evidență importanța și eficacitatea calculului matematic ca instrumentul auxiliar de analiză numerică în cadrul acestei cercetări.

Toate aceste probleme ca și rezolvarea lor teoretică și practică le presupunem cunoscute și de aceea nu le vom examina decît tangențial pentru a evita înțelegerea greșită, formală, ruptă de realitate a calculului matematic în statistică.

În acest sens, aplicînd procedeele matematice ce vor fi expuse în această parte ca și în celelalte părți ale lucrării, statisticianul economist nu va întîrzia să facă o prealabilă analiză politico-economică a fenomenelor și proceselor studiate. Numai o asemenea analiză, efectuată cu respectarea riguroasă a tezelor materialismului istoric și ale economiei politice marxist-leniniste, îi va oferi posibilitatea de a releva calitatea, esența fenomenelor studiate și, în consecință, de a fundamenta științific, aplicarea concretă atît a metodelor specifice statisticii, printre care metoda grupării ocupă primul loc, cît și a calculului matematic ca instrument auxiliar al statisticii, de netăgăduită utilitate.

*

Trecînd acum direct la expunerea elementelor de analiză numerică a distribuțiilor statistice, ne vom aminti, în primul rînd, că fiecare unitate a colectivității statistice studiate are anumite trăsături care se regăsesc în întreaga colectivitate. Aceste trăsături, semne sau proprietăți, se numesc în statistică *caracteristici*. Observării statistice îi sînt supuse, după cum știm, caracteristicile care *variază* de la unitate la unitate, prezentînd totodată calitatea de a fi *esențiale*, *tipice* pentru fenomenul supus cercetării statistice. Nu este exclusă și posibilitatea ca la una sau mai multe unități

ale colectivității totale, nivelul caracteristicii studiate să fie același. În majoritatea cazurilor, însă, nivelurile caracteristicii variază și tocmai această latură a fenomenelor constituie obiectul de cercetare a statisticii care folosește în acest scop și calculul matematic.

Dacă o caracteristică oarecare a unităților care formează colectivitatea totală se schimbă *cantitativ*, o vom numi *caracteristică variabilă*, iar mărimile ei concrete la fiecare unitate în parte, *variante*. De pildă, numărul aparatelor la care lucrează un muncitor într-o întreprindere automatizată, nivelul beneficiului realizat de diferite întreprinderi, productivitatea muncii etc. variază de la o unitate de observare la alta.

Exemple de fenomene ale căror manifestări pot fi caracterizate prin caracteristici variabile ni le pot oferi toate ramurile de activitate social-economică ale economiei naționale.

Variația caracteristicii se exprimă, după caz, prin numere raționale (ca în exemplele enumerate mai sus) sau prin numerele reale cuprinse într-un interval; în funcție de aceasta se face și împărțirea caracteristicilor variabile în caracteristici *discrete* și caracteristici *continue*.

Schimbarea caracteristicilor discrete se manifestă prin salturi; nivelul lor nu poate lua decît anumite valori pe scara lor de variație impusă de esența fenomenului. Deosebit de ilustrativ pentru cazul unei caracteristici discrete este exemplul clasic al variației numărului de copii de la o familie la alta. Acest număr nu poate lua decît valorile 0, 1, 2, 3,..., fracțiuni de copii, evident, nu va avea nici o familie.

Caracteristicile variabile continue pot lua *orice valori* pe scara de variație a fenomenului. Astfel de caracteristici sînt, de exemplu, forța motrice a unei instalații, producția globală a unei întreprinderi etc.

Trebuie să precizăm că, în practică, continuitatea caracteristicilor variabile este estompată de posibilitățile de exactitate reduse cu care putem efectua anumite măsurători. Pe de altă parte, caracteristicile variabilei continue apar uneori ca variabile discontinue, deoarece culegerea datelor privind mărimea diferitelor niveluri ale caracteristicii se face cu o aproximație mai mult decît grosolană.

Pe de altă parte, din punct de vedere practic, multe din variabilele discontinue pot fi considerate ca niște variabile continue. Așa, de pildă, populația unei țări are indiscutabil caracterul unei variabile discontinue și totuși salturile discontinuității sînt așa de mici față de ordinul de mărime al variabilei încît aceasta poate fi considerată ca o variabilă continuă, fără ca prin aceasta să se facă o eroare sensibilă.

Dacă vom despărți colectivitatea totală în grupe de unități cu același nivel al caracteristicii, vom determina în acest mod *frecvența* unor asemenea grupe care poate fi exprimată în valori absolute sau relative, dacă volumul întregii colectivități va fi considerat egal cu unitatea sau cu 100.

Seria numerelor absolute sau relative, puse față în față cu variantele corespunzătoare ale caracteristicii studiate, formează așa-numita funcție de *distribuție* a caracteristicii variabile considerate, în cadrul colectivității date.

Acestei funcții experimentale (empirice) i se mai spune și *serie de distribuție* (variație) sau pe scurt — distribuție statistică.

2. Funcție teoretică. Funcție experimentală

Noțiunea de funcție experimentală (empirică) sau funcție statistică nu trebuie confundată cu *funcția teoretică*, care exprimă legătura pur matematică dintre două mulțimi de obiecte.

Fiind date două mulțimi de obiecte a și A, între ele poate fi caracterizată o anumită dependență în felul următor:

Să notăm cu x obiectul generic sau punctul generic din mulțimea a. Presupunem că acest punct poate fi individualizat într-un mod oarecare astfel încît să aibă sens a se vorbi despre un anumit punct x; la fel presupunem că și în mulțimea A poate fi individualizat un anumit punct y. Dacă fiecărui punct x din a îi corespunde un anumit punct determinat y din A, corespondența aceasta $x < y$ este o *funcție* care se scrie

$$y = f(x).$$

Să considerăm, de pildă, un corp care cade; el face să se stabilească o corespondență între diferitele momente ale timpului $t(\equiv x)$ și pozițiile y pe care le ocupă în spațiu. Corespondența $y = f(t)$ este o funcție și reprezintă o lege naturală. Dacă avem de-a face cu o cădere pe verticală, fiecare poziție y a corpului în cădere poate fi caracterizată printr-un singur număr real, după cum și momentele t sînt caracterizate, de asemenea, prin numere reale. Corespondența este o funcție reală. Dacă avem de-a face cu un proiectil a cărui mișcare este curbilinie, fiecare poziție a lui este un vector în spațiul cu trei dimensiuni, în acest caz funcția $y = f(t)$ este vectorială.

Caracterizarea de mai sus este matematică. Corespondența $y = f(x)$ are caracter pur matematic în sensul că *y se consideră determinat o dată cu x.*

Dacă scriem de pildă, pentru legea căderii corpurilor în vid,

$$z = -\frac{1}{2}gt^2 + \left(\frac{dz}{dt}\right)_{t=0} t + z_0,$$

această funcție specială este expresia unei legi naturale după care corpurile în cădere liberă au accelerație constantă.

În general, o experiență duce la o funcție empirică.

Dacă, de pildă, măsurăm înălțimile *efective*

$$Z_0, Z_1, Z_2, \ldots, Z_{10}$$

la care s-a găsit un corp în cădere la o succesiune de momente înregistrate cu un cronometru — să zicem de la 1 la 10 secunde — obținem o funcție experimentală (empirică), exprimînd corespondența între numerele 1, 2,...,10 și numerele $Z_1, Z_2, \ldots, Z_{10}$.

Numerele $Z_1, Z_2, \ldots, Z_{10}$ nu sînt aceleași în fiecare experiență. Curenții atmosferici, forma și poziția corpului etc. ne conduc întotdeauna la alte valori Z_i, apropiate desigur între ele, dar diferite.

Să mai considerăm următorul exemplu: efectuăm cinci serii de experiențe, fiecare serie constînd din aruncarea de 100 ori a unei monede.

Scanned with OKEN Scanner

Numărăm pentru fiecare serie numărul de apariție al unei aceleiași fețe și-l raportăm la o sută.

Să admitem că am obținut rezultatele următoare :

Nr. seriei	Raportul dintre numărul de apariție al unei fețe determinate și o sută pentru seria respectivă
1	0,51
2	0,47
3	0,53
4	0,54
5	0,49

Între mulțimile (1 ; 2 ; 3 ; 4 ; 5) și (0,51 ; 0,47 ; 0,53 ; 0,54 ; 0,49) s-a creat corespondența dată de tabloul de mai sus, în urma experienței pe care am executat-o. Sîntem iarăși în prezența unei funcții statistice. În general, rezultatele din coloana a doua vor diferi de la experiență la experiență.

Verificările experimentale ne dau, deci, funcții empirice, adică corespondența între un număr finit de elemente $x_1, x_2, \ldots, x_n$ și elementele $y_1, y_2, \ldots, y_n$, caracterizate fiecare prin variațiile accidentale indisolubil legate de operațiile practice.

Compararea funcțiilor empirice cu funcția teoretică — cunoscută sau numai ipotetic cunoscută — constituie unul din cele mai interesante aspecte ale cercetării statistice teoretice și practice.

În afară de funcțiile de punct $y = f(x)$, un rol deosebit de important în statistică îl joacă *funcțiile de mulțime*.

Lungimea unui segment, aria unui poligon etc. sînt funcții de mulțimea de puncte la care se referă. Am văzut în partea întîi a acestei lucrări că și *probabilitatea* unui eveniment (care constă, în general, dintr-o mulțime de evenimente elementare) este tot o funcție de mulțime, teoretică, bazată pe stabilitatea frecvențelor experimentale de masă care au loc în anumite condiții.

Funcțiile experimentale de mulțime au aspecte foarte variate. În cercetările de statistică social-economică, funcțiile experimentale sînt exprimate, după cum am văzut, prin frecvențe.

3. Diferite tipuri de distribuții statistice și elementele lor generale

Ținînd seama de cele spuse în paragraful precedent, putem preciza acum că rezultatele unei înregistrări statistice se prezintă întotdeauna sub forma seriilor de distribuție în care caracteristicii variabile ce se manifestă sub forma unui anumit atribut sau a unui nivel cantitativ oarecare, îi corspund anumite frecvențe. În mod analog unei funcții teoretice, o serie de distribuție statistică poate fi privită ca o exprimare a unei variabile în raport cu variația alteia ; ea se prezintă, deci, ca o corespondență între două șiruri de valori.

Diferențierea pe grupe a unei colectivități statistice, în funcție de variația caracteristicii și a frecvențelor corespunzătoare se poate face în primul rînd, după locul unde s-a produs fenomenul, în al doilea rînd, în raport cu timpul cînd s-a făcut observarea lui și, în sfîrșit, în raport cu un atribut oarecare, altul decît spațiul sau timpul.

În mod corespunzător se deosebesc trei tipuri de distribuții :

a) distribuții în spațiu,

b) distribuții în timp (serii dinamice) și

c) distribuții de frecvențe.

În această parte vom studia numai distribuțiile de frecvențe, după o prezentare sumară și a seriilor de distribuții în spațiu, urmînd ca seriilor în timp (dinamice) să le fie rezervată partea a III-a.

Seriile statistice în care elementul de diferențiere a manifestărilor fenomenului este locul în care aceste manifestări s-au produs se numesc *distribuții în spațiu.*

În tabelul 6 dăm un exemplu de trei distribuții în spațiu paralele.

Tabelul 6

Distribuția suprafeței R.P.R. ca și a numărului și densității populației ei pe regiuni, la 1 iulie 1961

Regiunea	Suprafața în km²	Numărul populației (locuitori)	Densitatea populației (loc/km²)
0	1	2	3
R.P.R.	237 500	18 566 932	78,2
Argeș	15 800	1 177 966	74,6
Bacău	13 400	1 072 315	80,0
Banat	21 800	1 234 342	56,6
Brașov	15 090	1 043 527	69,2
București	20 480	1 673 450	81,7
Cluj	16 820	1 207 372	71,8
Crișana	12 240	874 374	71,4
Dobrogea	15 460	507 862	32,9
Galați	12 910	1 044 656	80,9
Hunedoara	11 000	637 905	58,0
Iași	11 100	1 029 687	92,8
Maramureș	10 500	767 005	73,0
Mureș-Aut. Maghiară	12 250	806 130	65,8
Oltenia	20 300	1 561 953	76,9
Ploiești	13 100	1 440 094	109,9
Suceava	13 750	985 731	71,7
Orașul București	970	1 354 836	1 396,7
„ Constanța	530	147 727	278,7

Sursa : Breviar Statistic al R.P.R., 1962, pp. 18 și 24.

După cum se vede, distribuțiile în spațiu se prezintă sub forma unor șiruri de date a căror ordine logică poate fi adoptată în chip diferit. În tabelul 6, datele au fost înșiruite în ordinea alfabetică a regiunilor, după ce în primul rînd au fost prezentate date pe ansamblul țării ; șirurile se încheie cu date privind întinderi spațiale care au anumite particularități de organizare — orașul București și orașul Constanța.

Scanned with OKEN Scanner

De obicei, epoca la care s-a făcut înregistrarea datelor constituie o caracteristică a distribuției în spațiu și de aceea trebuie menționată o dată cu alte elemente privitoare la culegerea datelor.

Seriile statistice în care variabila independentă este reprezentată de o caracteristică oarecare, alta decît spațiul sau timpul, se numesc *distribuții de frecvențe*.

Caracteristica în raport cu care se stabilește distribuția de frecvență poate fi *calitativă* sau *cantitativă* *).

De exemplu, distribuția populației în raport cu starea civilă, sau pe sexe este o distribuție după o caracteristică calitativă; starea civilă sau sexul nu pot fi exprimate numeric decît, eventual, în mod convențional.

Tabelul 7 ne oferă un exemplu de distribuții de frecvență după o caracteristică calitativă.

Tabelul 7

Distribuția populației orașului București după starea civilă și pe sexe
(Recensămîntul populației din 21.II.1956)

Starea civilă	Total populație	din care:	
		masculin	feminin
0	1	2	3
Total	1 236 908	591 461	645 447
Necăsătoriți	468 947	247 453	221 494
Căsătoriți	647 095	325 604	321 491
Divorțați	37 827	8 773	29 054
Văduvi	81 190	8 985	72 205
Nedeclarată	1 849	646	1 203

Sursa: *Recensămîntul populației din 21 februarie 1956.* Rezultate generale D.C.S.

Este de observat că distribuțiile în spațiu se apropie foarte mult, prin forma sub care se prezintă, de distribuțiile de frecvență în raport cu o caracteristică calitativă.

În tabelul 8 este dată o distribuție după o caracteristică cantitativă discontinuă.

Din acest tabel se observă că întreaga colectivitate de familii cu trei persoane a fost împărțită în patru grupe după o caracteristică comună și anume: numărul persoanelor active din familie; această caracteristică — — variabila independentă — a luat patru valori întregi (de la 0 la 3) cărora le-au corespuns frecvențe de grupă bine determinate (coloana 1). Numărul unităților unei grupe poartă numele *de frecvență absolută* a acelei grupe din cadrul colectivității totale, iar numărul tuturor elementelor colectivității, poartă numele de *volum* al colectivității.

*) Cunoscutul statistician suedez C. Charlier a folosit denumirile de distribuții *homograde* și distribuții *heterograde*, pentru distribuții de frecvență după o caracteristică calitativă și respectiv, după caracteristica cantitativă. În prezent această terminologie nu are decît o circulație foarte limitată.

Tabelul 8

Distribuția numărului de familii din R.P.R. cu trei persoane după numărul persoanelor active din familie

Numărul persoanelor active din familie x	Numărul familiilor alcătuite din trei persoane F	f %
0	1	2
0	40 542	3,63
1	30 5648	27,34
2	516 161	46,17
3	255 587	22,86
Total	1 117 948	100,00

Sursa : *Recensămîntul populației din 21 februarie 1956.* Date generale D.C.S.

Volumul total al colectivității ilustrate în tabelul 8 este de 1 117 948 familii de cîte trei persoane.

Se mai observă că în tabel (coloana 2), frecvențele absolute ale fiecărei grupe au fost raportate la volumul total al colectivității, obținîndu-se un alt șir de frecvențe care poartă numele de *frecvențe relative*. Calculul acestor frecvențe se face de obicei prin proporția de 1, 10, 100 etc.; în tabel ele s-au calculat în procente (%).

Notînd cu F_i frecvența absolută a unei grupe, cu $\sum F_i$ suma frecvențelor (volumul colectivității totale), frecvența relativă este dată de formula

$$f = p\,\frac{F_i}{\sum F_i}, \qquad (p = 1,\ 10,\ 100,\ldots).$$

Prin caracterul caracteristicii independente, distribuția din tabelul 8 este o distribuție *discretă*.

O altă formă de prezentare o are distribuția de frecvență din tabelul 9.

Distribuția din acest tabel are particularitatea că variabila independentă — vîrsta — este măsurată în intervale egale de cîte 5 ani (0—4, 5—9, 10—14,..., 95—99 și 100 ani și peste) și că ea poate lua orice valoare cuprinsă în scara ei firească de variație, între și 0 și peste 100 ani. Avem prin urmare de-a face cu o distribuție *continuă* în care frecvențele absolute (coloanele 1, 2 și 3), $F(x)$ reprezintă o funcție de *intervale* de valori, menținîndu-se corespondența între valorile de interval ale caracteristicii variabile și frecvențele respective. În coloanele 2 și 3 sînt date frecvențele absolute ale populației grupate pe cele două sexe, păstrîndu-se totodată corespondența cu intervalele caracteristicii variabile independente. Asupra acestei particularități vom mai reveni.

Să considerăm încă două exemple (tabelele 10 și 11) de distribuții continue de intervale.

Și în aceste tabele sînt ilustrate serii de variație de intervale. Observăm că în tabelul 11, intervalele sînt inegale și că în el se combină două serii de variație paralele : variația numărului de localități și a numărului

Tabelul 9

Distribuția populației R.P.R. pe grupe de vîrstă

(Recensămîntul populației din 21 februarie 1956).

Grupa de vîrstă x	Total R.P.R. F	din care : masculin F_m	din care : feminin F_f
0	1	2	3
Total populație	17 489 450	8 503 420	8 986 030
0— 4 ani	1 848 812	944 338	904 474
5— 9 „	1 634 885	831 840	803 045
10—14 „	1 329 917	675 183	654 734
15—19 „	1 582 215	782 125	800 090
20—24 „	1 595 379	805 084	790 295
25—29 „	1 560 772	785 523	775 249
30—34 „	1 374 768	660 178	714 590
35—39 „	856 488	389 721	466 767
40—44 „	1 130 498	524 938	605 560
45—49 „	1 110 829	543 187	567 642
50—54 „	928 660	458 910	469 750
55—59 „	800 037	373 984	426 053
60—64 „	620 898	268 861	352 037
65—69 „	480 119	201 911	278 208
70—74 „	346 771	145 175	201 596
75—79 „	180 989	71 461	109 528
80—84 „	72 140	28 348	43 792
85—89 „	21 186	9 176	16 010
90—94 „	6 212	1 963	4 249
95—99 „	1 913	486	1 427
100 ani și peste	487	103	384
Nedeclarată	1 475	925	550

Tabelul 10

Distribuția stațiunilor de mașini și tractoare după volumul mediu de lucrări pe un tractor *) (în ha arătura normală) la sfîrșitul anului 1959

Grupă a volumului mediu de lucrări pe un tractor	Unități F_i	% f
0	1	2
Total	242	100,0
Pînă la 250 ha	18	7,4
De la 251—300 ha	58	24,0
„ „ 301—350 „	47	19,4
„ „ 351—400 „	51	21,1
„ „ 401—450 „	36	14,9
Peste 450 „	32	13,2

*) Tractor mediu convențional de 15 C.P. (scriptic).

Sursa : *Breviar statistic al R.P.R.*, 1960, p. 131.

Tabelul 11

Distribuția localităților urbane și a populației urbane după numărul de locuitori la 1 iulie 1959

Grupă a numărului de locuitori x	Numărul localităților F_i	Numărul populației F_i	% față de totalul populației $f = \frac{F_i}{\Sigma F_i}$
0	1	2	3
Total urban R.P.R. *)	354	5 838 401	100,0
Sub 3 000 locuitori	60	126 492	2,2
3 000 — 4 999 locuitori	85	331 075	5,7
5 000 — 9 999 „	97	673 294	11,5
10 000 — 24 999 „	75	1 154 620	19,8
25 000 — 49 999 „	20	662 295	11,3
50 000 — 99 999 „	5	333 193	5,7
100 000 — și peste — „	11	1 330 321	22,8
Orașul București	1	1 227 111	21,0

*) Inclusiv 183 localități asimilate mediului urban.
Sursa: *Breviar statistic al R.P.R.*, 1960.

populației, ambele în raport cu aceeași caracteristică — numărul de locuitori în diferite localități.

Trei probleme deosebit de importante se ridică în legătură cu elaborarea seriilor de variație și anume: problema alegerii intervalelor scării de variație, problema limitei intervalelor de grupare și problema intervalelor marginale (închise și deschise). Le vom examina pe scurt.

A. Problema alegerii intervalelor seriei de variație. Modul cum se aleg intervalele care formează scara de grupare a unităților prezintă o importanță deosebită în studiul seriilor de variație. De el depinde, în ultima analiză, caracterul grupării, posibilitatea acesteia de a releva pe baza analizei calitative, tipurile esențiale ale distribuției, specificul ei.

Am văzut că în seriile ilustrate în tabelele 9 și 10, intervalele de grupare erau egale, pe cînd în tabelul 11 ele variază după poziția lor în scară. Pentru a face posibilă compararea directă a frecvențelor de grupă este recomandabil ca, în general, orice serie de variație să fie construită pe o scară cu intervale egale. Necesitatea acestei cerințe derivă nemijlocit din împrejurarea că tocmai diferența de mărime dintre frecvențele de grupă constituie elementul esențial, cel mai caracteristic al unei distribuții.

De pildă, în cazul tabelului 11, frecvența corespunzătoare intervalului 3 000 — 4 999 locuitori nu este comparabilă cu frecvența grupei imediat următoare 5 000 — 9 999. Între prima frecvență în valoare de 85 și a doua, în valoare de 97 localități sau, respectiv, de 331 075 și 673 294 locuitori (dacă ne referim la funcția empircă care exprimă numărul populației), are loc un salt datorit faptului că al doilea interval considerat este de 2,5 ori mai mic decît intervalul al treilea (tabelul 11). Cu toate acestea, adoptarea intervalelor de grupare inegale se impune uneori cu necesitate,

aceasta îndeosebi în cazul unei mari amplitudini de variație a caracteristicii de grupare care — în ipoteza unor intervale egale numeroase — ar face să apară o distribuție cu multe frecvențe de grupă nule.

De acest lucru ne putem da seama cu ușurință dacă vom regrupa, de pildă, datele din tabelul 11 pe baza unor intervale egale, să zicem de 50 000 locuitori. Distribuția nouă (considerînd numai distribuția localităților) va arăta ca în tabelul 12.

Tabelul 12

Distribuția localităților urbane după numărul de locuitori la 1 iulie 1959

Grupă a numărului de locuitori x	Numărul localităților F_i
0	1
Total R.P.R. *)	354
Sub 50 000 locuitori	337
50 000— 99 999 locuitori	5
100 000—149 999 ,,	10
150 000—199 999 ,,	1**)
200 000—249 999 ,,	0
250 000 — 299 999 ,,	0
Peste 1 200 000 ,,	1

*) Incluse 183 localități asimilate mediului urban.

**) Frecvența se referă la orașul Cluj, care are 162 419 locuitori.

După cum se vede, această nouă distribuție nu ne spune altceva decît că aproape totalitatea localităților urbane sau asimilate lor (96,6%) au o populație mai mică decît 100 000 de locuitori, ceea ce înseamnă că valoarea cognitivă a distribuției întocmite este departe de a fi satisfăcătoare. Forma distribuției din tabelul 12 are și dezavantajul de a prezenta numeroase grupe (20), lipsite de orice interes deoarece au o frecvență nulă.

În afară de aceasta, frecvențele din unele grupe apar sub forma unei simple enumerări, aceasta datorită poziției singulare a unor unități, în ansamblul localităților urbane (Cluj, București).

Aceste inconveniente se pot evita întocmindu-se o grupare cu intervale mai mari către partea superioară a distribuției, ca în tabelul 11. Se ajunge astfel — în mod firesc — la necesitatea fixării unor intervale inegale.

Stabilirea mărimii intervalului de grupare este o chestiune de cunoaștere profundă a esenței fenomenului studiat, ce nu se poate realiza decît aplicînd principiile analizei calitative marxist-leniniste.

Se știe, că în general, o grupare efectuată pe bază de intervale mici poate scoate în evidență lucruri de amănunt, particulare, care nu aparțin legității fenomenului de masă luat în studiu și invers, o grupare efectuată pe bază de intervale mari, atenuează iregularitățile distribuției, făcînd

să se manifeste legăturile legice, obiectiv determinate, *esențiale* și *necesare*, în întregul ansamblu al fenomenului.

Este totuși cazul să atragem atenția că orice statistician trebuie să aibă grija deosebită de a nu alege nici intervale prea mari, străine de natura fenomenului studiat, pentru a nu ajunge la efectul contrar celui dorit și anume, la estomparea totală sau parțială a trăsăturilor aceleași legități statistice.

În legătură cu mărimea și conținutul intervalelor se ivește de multe ori necesitatea ca intervalele grupelor marginale să fie subgrupate pe bază de fracțiuni simple din mărimea comună a celorlalte. Această necesitate se vădește din cauza faptului că unitățile care formează grupele marginale au un caracter excepțional, constituind partea delicată a unei distribuții.

O asemenea metodologie s-a adoptat, de obicei, la întocmirea distribuției populației unei țări pe grupe de vîrstă, subdivizîndu-se, în intervale egale, grupa inițială a copiilor de sub un an pentru a se putea studia mortalitatea infantilă. O astfel de metodologie a fost adoptată și la gruparea populației din R.P.R. ,la recensămîntul din 21 februarie 1956.

Rezolvînd problema mărimii intervalului sau — ceea ce este același lucru — a numărului de grupe, trebuie să avem în vedere o considerație importantă și anume : numărul grupelor trebuie stabilit în așa fel încît în cadrul fiecărei grupe, să nu avem abateri mari de la o distribuție simetrică. Acest procedeu se dovedește a fi foarte util mai tîrziu la interpretarea seriei de intervale ca și la prelucrările ei ulterioare cînd valoarea de la mijlocul fiecărui interval trebuie să ne servească drept mijloc de apreciere a mărimii celorlalte valori din cadrul grupei. Dacă, de exemplu, ne referim la grupa de vîrstă 45—49 ani din tabelul 9, putem considera că valoarea de 47 ani este caracteristică pentru toate unitățile (în număr de 1 110 829) care fac parte din această grupă. Desigur, o asemenea prezumție nu se verifică cu precizie decît foarte rar și de obicei atunci cînd valoarea de la mjilocul intervalului coincide cu media aritmetică a tuturor valorilor din interval sau are o abatere nesemnificativă. Această cerință s-ar realiza aproximativ în cazul cînd distribuția s-ar împărți în numeroase intervale mici. Or — după cum știm — a proceda astfel ar însemna să neglijăm necesitatea de a releva regularitatea de ansamblu a distribuției.

B. *Problema limitei intervalelor de grupare.* În practica statisticii au fost experimentate diferite valori de precizare a limitelor de intervale ale unei serii de variație.

Forma cea mai des întîlnită este de tipul (x) de mai jos :

primul interval	0—1	
al doilea interval	1—2	(α)
al treilea interval	2—3	
al patrulea interval	3—4	
. . . . .		

în care limita inferioară a intervalului se consideră inclusă în interval. De exemplu, valoarea 2 care poate face parte din intervalul al doilea ca și din intervalul al treilea se consideră, neîndoielnic, că aparține intervalului al treilea.

Scanned with OKEN Scanner

O altă formă simplă ce se deduce din forma (α) este următoarea:

$$\begin{matrix} 0- \\ 1- \\ 2- \\ 3- \\ \vdots \end{matrix} \qquad (\alpha')$$

în care limita superioară a intervalului este suprimată, fiind egală cu limita intervalului imediat următor de pe scara de variație.

O formă ce se întîlnește adeseori în lucrările statistice este și aceea de tipul (β) de mai jos:

primul interval	0—0,9	(β)
al doilea interval	1—1,9	
al treilea intreval	2—2,9	
al patrulea interval	3—3,9	
. . . . .		

Această formă trebuie interpretată în sensul că: primul interval va cuprinde toate valorile care au la partea întreagă cifra 0, iar la prima zecimală oricare din cifre, pînă la 9 inclusiv; al doilea interval va cuprinde valorile ce au la partea întreagă 1 etc. De exemplu, valoarea 0,97 va fi cuprinsă în primul interval.

Forma cea mai explicită ar fi de tipul (γ) de mai jos:

primul interval	mai mare sau egal cu 0 și mai mic decît 1,	
al doilea interval	„ „ „ „ „ 1 „ „ „ „ 2,	(γ)
al treilea interval	„ „ „ „ „ 2 „ „ „ „ 3,	
etc.		

Această formă are, însă, dezavantajul de a ocupa un spațiu prea mare și de aceea nu este folosită în practică.

C. Problema intervalelor marginale — închise și deschise. Primul și ultimul interval al unei serii de variație au semnificații deosebite în statistică, reprezentînd cazuri excepționale de pe scara firească de variație a caracteristicii studiate, uneori cazuri cu totul particulare. Se impune de aceea ca aceste intervale să fie *închise*, precizîndu-se concret limitele lor, inferioară și superioară, sau cel puțin limitele uneia din ele. Adeseori, însă, această cerință nu poate fi îndeplinită. Așa, de pildă, a fost cazul distribuției populației pe grupe de vîrstă (vezi tabelul 9) în care primul interval este *închis* iar ultimul — abstracție făcînd de grupa celor care n-au declarat vîrsta — este deschis. Nu trebuie să se confunde noțiunile de interval deschis și interval închis din analiza matematică cu aceste noțiuni, care sînt intrate de multă vreme în terminologia statistică.

Intervale marginale deschise au și exemplele de distribuții date în tabelele 10 și 11.

În general, lipsa de precizie care derivă din forma deschisă a unei serii de variație o face să fie mai greu utilizabilă. În cazul cînd intervalul deschis este la partea superioară a scării, neajunsul seriei este mai vădit și acesta este tocmai cazul care se întîlnește cel mai des în practică. Analiza unor asemenea serii necesită, de obicei, informații suplimentare relative la distribuția reală din cadrul intervalelor deschise; pentru efec-

tuarea ei în condiții optime, statisticianul trebuie să manifeste o grijă și o prudență deosebită, ferindu-se îndeosebi, de extrapolări care pot duce la rezultate cu totul greșite.

4. Densitatea distribuției

A determina densitatea distribuției pe grupe înseamnă a stabili care anume variantă a caracteristicii variabile are frecvența relativă cea mai mare. Este evident că această problemă se rezolvă cu ușurință cînd avem de-a face cu o distribuție discretă, ca în tabelul 8, în care este ilustrată distribuția numărului de familii din R.P.R. — alcătuite din trei persoane după numărul persoanelor active din familie. Se vede că familiile cu două persoane active (din trei) reprezintă grupa cea mai numeroasă, 46,17% din totalul familiilor considerate.

Problema se complică, însă, cînd trebuie să stabilim densitatea unei distribuții continue sau a unei distribuții discrete care, practic, poate fi considerată continuă (am arătat că din această categorie fac parte funcțiile experimentale discontinue la care variația se face în salturi așa de mici încît valorile respective nu se pot supune unei măsurări sau numărări precise și — după cum vom vedea — nici unei reprezentări grafice perceptibile).

Să considerăm, de exemplu, distribuția populației R.P.R. pe grupe de vîrstă (pe baza datelor recensămîntului populației din 21 februarie 1956; vezi și tabelul 9), în care se scoate în evidență frecvența relativă și densitatea distribuției (tabelul 13).

Tabelul 13

Grupa de vîrstă x	Frecvența relativă (%) f	Densitatea distribuției $f(x)$
0	1	2
TOTAL R.P.R.	100,0	
0— 4 ani	10,6	2,10
5— 9 „	9,3	1,86
10—14 „	7,6	1,52
15—19 „	9,0	1,80
20—24 „	9,1	1,82
25—29 „	8,9	1,78
30—34 „	7,9	1,58
35—39 „	4,9	0,98
40—44 „	6,5	1,30
45—49 „	6,4	1,28
50—54 „	5,3	1,06
55—59 „	4,6	0,92
60—64 „	3,6	0,72
65—69 „	2,7	0,54
70—74 „	2,0	0,40
75—79 „	1,0	0,20
peste 80 „	0,6	0,12

Scanned with OKEN Scanner

Acest tabel creează impresia că cel mai frecvent se întîlnește populația sub vîrstă de 5 ani.

Dacă întrunim însă grupele a treia și a patra în una singură și apoi a cincea și a șasea ș.a.m.d., vom obține distribuția din tabelul 14 a.

Tabelul 14 a

Grupa de vîrstă x	Frecvența relativă (%) f	Densitatea distribuției $f(x)$
0	1	2
Total R.P.R.	100,0	
0— 4 ani	10,6	2,12
5— 9 „	9,3	1,86
10—19 „	16,6	1,66
20—29 „	18,0	1,80
30—39 „	12,8	1,28
40—49 „	12,9	1,29
50—59 „	9,9	0,99
60—69 „	6,3	0,63
70—79 „	3,0	0,30
peste 80 „	0,6	0,06

Acum pare că cea mai des întîlnită populație este cea de vîrstă cuprinsă între 20 și 30 ani, ea reprezentînd 18% din totalul populației.

În sfîrșit, dacă întrunim și primele două grupe în una singură, obținem o altă distribuție (tabelul 14 b).

Tabelul 14 b

Grupa de vîrstă x	Frecvența relativă (%) f	Densitatea distribuției $f(x)$
0	1	2
0— 9 ani	19,9	1,99
10—19 „	16,6	1,66
20—29 „	18,0	1,80
30—39 „	12,8	1,28
40—49 „	12,9	1,29
50—59 „	9,9	0,99
60—69 „	6,3	0,63
70—79 „	3,0	0,30
peste 80 „	0,6	0,06

La transformările de mai sus ale distribuției inițiale, sporirea frecvenței relative în diferitele grupe ale colectivității s-a obținut prin formarea unor intervale mai mari, întrunindu-se grupele vecine într-una singură. Se vede astfel că frecvența relativă corespunzătoare unui interval oare-

Scanned with OKEN Scanner

care depinde de mărimea acestuia. Dacă, de pildă, întrunim două intervale vecine de mărimi egale, cu aceeași frecvență, obținem un interval mai mare cu frecvență dublă.

Prin urmare, măsura densității unei variante a caracteristicii continue (sau practic-continue), cuprinsă într-un interval oarecare, nu poate fi însăși frecvența, ci *raportul ei față de mărimea intervalului.*

Acest raport se numește *densitatea distribuției.* Am văzut în partea întîi la capitolul VII că, pe plan probabilistic, noi am considerat noțiunea teoretică de densitate de probabilitate în punctul x pe care am notat-o cu $f(x)$ și care, după cum am văzut, este derivata funcției de repartiție continue $F(x)$*). Tocmai acestei noțiuni teoretice îi corespunde, în cadrul funcțiilor experimentale continue, noțiunea de densitate a distribuției.

Vom face aici cîteva precizări în legătură cu exemplul considerat mai sus.

În ultima coloană a tabelelor 13, 14 a și 14 b este dată densitatea distribuției. Compararea acestor tabele învederează faptul că nu există nici o contradicție între datele pe care le cuprind ele. Astfel, ultimul tabel arată că densitatea cea mai mare a distribuției a fost în intervalul de la 0 la 10 ani, iar primul tabel arată aceasta cu mai mare precizie, relevînd că densitatea distribuției este mai mare în prima parte a acestui interval și anume la populația de vîrstă sub 5 ani.

Presupunînd că în scopul unei precizii mai mari în ilustrarea distribuției avem posibilitatea de a trece la intervale tot mai mici, calculîndu-se de fiecare dată și densitatea distribuției pentru fiecare interval, vom obține valori ale acestora din urmă tot mai mici, deoarece am presupus distribuția continuă.

Dacă în acest caz notăm cu M_a^b frecvența relativă (în care a și b simbolizează limitele inferioară și superioară ale intervalului), în intervalul $(x, x + \Delta x)$ densitatea distribuției va fi dată de raportul

$$\frac{M_x^{x+dx}}{\Delta x}. \qquad \text{(IX,4.1)}$$

Dacă pentru $\Delta x \to 0$, densitatea distribuției tinde către o limită oarecare, această limită poate fi considerată ca fiind densitatea distribuției „în punctul dat", corespunzător valorii lui x în acel punct și exprimată prin funcția

$$f(x) = \lim_{\Delta x \to 0} \frac{M_x^{x+\Delta x}}{\Delta x}. \qquad \text{(IX,4.2)}$$

Trebuie să precizăm că una din problemele cele mai importante ce se pune statisticianului în cadrul procesului de analiză cantitativă a unei distribuții statistice este tocmai aceea de a o caracteriza sintetic printr-o funcție oarecare $f(x)$ ce măsoară densitatea ei.

*) Vezi explicațiile date la pag. 178.

5. Frecvența cumulată — integrală a densității distribuției

Numărătorul formulei (IX,4.1), prin care se calculează densitatea distribuției, capătă o interpretare mai simplă, introducînd noțiunea de *frecvență cumulată* sau *ogiva* pe care o vom nota cu $F(x)$. Frecvența cumulată, corespunzînd unei valori a caracteristicii independente x, reprezintă acea parte a volumului colectivității studiate ale cărei unități sînt purtătoare ale variantelor caracteristicii mai mici decît x. Folosind notațiile din paragraful precedent, putem scrie

$$F(x) = M_{-\infty}^{x}$$

ținînd seama că unitățile colectivității care au un nivel al caracteristicii egal cu limita superioară a intervalului *sînt incluse* în interval, iar cele cu un nivel al caracteristicii egal cu limita lui inferioară *sînt excluse*.

După cum vedem, noțiunii de frecvență cumulată îi corespunde — pe plan teoretic, probabilistic — noțiunea de *funcție de repartiție* pe care am examinat-o tot în capitolul VI.

Dăm în tabelul 15 frecvența cumulată a distribuției populației R.P.R., pe baza datelor recensămîntului din 1956.

Tabelul 15

Vîrsta	Frecvența relativă cumulată (%)
A	1
0 ani	0
10 „	19,9
20 „	36,5
30 „	54,5
40 „	67,2
50 „	80,1
60 „	90,0
70 „	96,3
80 „	99,3
	100,0

După cum știm încă din capitolul VI, în cazul unei caracteristici continue avem

$$M_a^b = F(b) - F(a)$$

și în consecință

$$f(x) = \lim_{\Delta x \to 0} \frac{F(x + \Delta x) - F(x)}{\Delta x}. \qquad \text{(IX,5.1)}$$

Dacă o asemenea limită există este ușor de văzut că ea este derivata frecvenței cumulate :

$$f(x) = \frac{\mathrm{d}F(x)}{\mathrm{d}x},$$

de unde

$$\mathrm{d}F(x) = f(x)\,\mathrm{d}x$$

și

$$F_a^b = F(b) - F(a) = \int_a^b f(x)\,\mathrm{d}x. \qquad \text{(IX,5.2)}$$

Am ajuns, deci, la concluzia că *frecvența corespunzătoare unui interval de variație de la a la b a caracteristicii studiate este integrala definită a densității distribuției, în limitele intervalului considerat.*

Această definiție nu necesită nici un fel de alte precizări dacă avem de-a face cu o caracteristică continuă și o colectivitate statistică *infinită* sau practic infinită. În caz contrar, relația (IX,5.2) nu-și poate găsi aplicație directă.

Trebuie să remarcăm că după ce frecvența relativă este exprimată prin integrala densității, raportul :

$$\frac{\int_a^b f(x)\,\mathrm{d}x}{b-a} \qquad \text{(IX,5.3)}$$

ne va da valoarea medie a funcției $f(x)$ în intervalul dat (a și b fiind limitele intervalului).

Așadar, densitățile distribuției calculate mai sus sînt valori medii.

În cazul unei distribuții statistice continue, știm din analiza matematică că există întotdeauna un număr ξ cuprins în intervalul (a,b), astfel ca

$$\frac{\int_a^b f(x)\,\mathrm{d}x}{b-a} = f(\xi).$$

Această egalitate exprimă o proprietate interesantă a densității unei distribuții statistice continue și anume că pentru orice interval finit există un punct al intervalului în care $f(\xi)$ să ne exprime în mod exact densitatea medie a intervalului ; ξ poate fi considerat din acest punct de vedere o medie a intervalului.

Este evident că dacă $f(x)$ aparține unei clase particulare de funcții, intervalul de variație a mediei poate deveni mai restrîns decît intervalul (a,b). D. Pompeiu *) a arătat că dacă $f(x)$ este un polinom de grad n, intervalul de variație al lui ξ este mai restrîns, fiind determinat de cea mai mică și cea mai mare rădăcină a unor polinoame ortogonale, de tip Legendre, definite pe acest interval.

*) D. Pompeiu (1873—1954), matematician romîn, a publicat lucrări importante în teoria funcțiilor de o variabilă complexă. A introdus în matematică noțiunea de derivată areolară.

Dacă luăm primele două grupe din tabelul 14 b și calculăm densitatea de distribuție medie, vom avea:

Vîrstă	Densitatea
0— 9 ani	1,99
10—19 „	1,66
0—19 „	1,825

$$1{,}825 = \frac{(1{,}99 \cdot 10) + (1{,}66.10)}{20}$$

adică, densitatea intervalului întrunit este dat de media aritmetică ponderată a densităților fiecărui interval în parte, în care ponderile sînt intervalele, măsurate în unitățile de măsură respective.

6. Descompunerea intervalului și determinarea frecvențelor parțiale corespunzătoare

O problemă de importanță practică deosebită constă în a se determina frecvența relativă corespunzătoare unor subgrupe ale intervalului dintr-o serie de variație, cunoscînd frecvența relativă a întregului interval considerat și frecvențele intervalelor precedent și următor.

E de observat că pentru a avea serii de variație cu intervale egale și inegale, tehnica de calcul diferă de la un caz la altul, fiind mai puțin complicată cînd avem intervale egale. Să ne oprim asupra fiecărui caz în parte.

1°. Să considerăm exemplul distribuției din tabelul 16 în care un număr de 86 sovhozuri din două regiuni din U.R.S.S. se distribuie după volumul producției globale ce a revenit pe cap de lucrător în 1958 (producția s-a calculat în prețuri de livrare).

Tabelul 16

Distribuția a 86 sovhozuri din două regiuni ale U.R.S.S., după volumul producției globale ce a revenit pe un lucrător *) (1958)

Grupe de sovhozuri după volumul producției globale pe un lucrător x	Numărul sovhozurilor % f
TOTAL	100,0
Sub 15 mii ruble	10,47
15,1—25 mii ruble	34,88
25,1—35 „ „	25,58
35,1—45 „ „	15,12
Peste 45 „ „	13,95

*) După datele publicate în revista „Vestnik Statistiki", 9 (1960). (Studiul lui B. Braghinski, tabel 4, p. 6).

Pe baza datelor din tabelul 16 ne putem pune întrebarea care anume este procentul sovhozurilor în care volumul producției globale ce revine pe un lucrător variază între 25,1 și 30 mii ruble (această variație este în cadrul intervalului grupei a 3-a).

A răspunde direct la această întrebare nu este posibil deoarece nu cunoaștem care este procentul (din 25,58%) al sovhozurilor la care volumul producției ce revine pe cap de lucrător este cuprins în intervalul de 30—35 mii ruble.

Problema poate fi rezolvată cu un grad de precizie suficient pentru nevoile practice în felul următor *)

Folosim notațiile:

A frecvența relativă a intervalului precedent celui considerat (34,88%);

B frecvența relativă a intervalului considerat (25,58%);

C frecvența relativă a intervalui următor (15,12%);

$\left.\begin{aligned} B - A &= \Delta F_{-1} \\ C - B &= \Delta F_0 \end{aligned}\right\}$ sporurile frecvenței relative;

$\Delta^2 F = (C-B) - (B-A)$ sporul frecvenței relative pe cele trei intervale;

$F_0^{0,5}$ = frecvența intervalului ce ne interesează;

θ mărimea intervalului 25,1—30, față de mărimea întregului interval (în cazul nostru $\theta = 0,5$).

Apoi, considerînd că densitatea distribuției $f(x)$ în intervalul 25,1—35 este suficient de apropiată de o parabolă de gradul doi (la fel și densitatea în intervalele vecine), se poate demonstra că

$$F_0^{0,5} = 0,5 \left[\frac{A + B + \Delta F_{-1} \cdot 0,5}{2} - \frac{\Delta^2 F (1 - 0,5^2)}{6} \right]$$

și cum

$$\Delta F_{-1} = 25,58 - 34,88 = -9,30$$

și

$$\Delta^2 F = (15,12 - 25,58) - (25,58 - 34,88) = -10,46 - (-9,30) = -1,16,$$

vom avea

$$F_0^{0,5} = 0,5 \left[\frac{34,88 + 25,58 + [(-9,30) \cdot 0,5]}{2} - \frac{-1,16 (1 - 0,5^2)}{6} \right] =$$

$$= 0,5 \left[\frac{34,88 + 25,58 - 4,65}{2} - \frac{-0,87}{6} \right] =$$

$$= 0,5 \, [27,9 - (-0,145)] = = 0,5 \cdot 28,045 =$$

$$= 14,0225 = 14,02\%.$$

*) Expunem aici metoda propusă de A. I. Boiarski, *Matematika dlia ekonomistov*. Moskva, 1957, p. 247—248.

Prin urmare, în aproximativ 14,02% din 25,58% sovhozuri, volumul producției globale ce revenea pe cap de lucrător în 1958 — în cele două regiuni ale U.R.S.S. — a variat între 25,1 și 30 mii ruble. Aceasta înseamnă aproximativ 50% din frecvența relativă a întregii grupe; după cum se vede, în interiorul intervalului 25,1—30, distribuția frecvențelor a fost aproximativ uniformă.

Demonstrația formulei folosite este simplă.

Am presupus că

$$f(x) = a_0 + a_1 x + a_2 x^2$$

în care valoriile x își au originea în limita inferioară a intervalului pe care vrem să-l descompunem (25,1 = 0). Considerînd această condiție în intervalul supus descompunerii, x va lua valorile cuprinse între 0 și 1, în intervalul precedent de la — 1 la 0, iar în intervalul următor, de la 1 la 2.

Pentru a determina cei trei parametri pentru $f(x)$, vom folosi frecvențele relative din cele trei intervale care ne sînt cunoscute.

După cum știm, frecvența relativă a unui interval este dată de integrala densității distribuției în limita intervalului respectiv.

Pentru cele trei intervale vom avea:

$$\int_{-1}^{0} f(x)\mathrm{d}x = \int_{-1}^{0} (a_0 + a_1 x + a_2 x^2)\,\mathrm{d}x =$$

$$= \left[a_0 x + \frac{a_1}{2} x^2 + \frac{a_2}{3} x^3\right]_{-1}^{0} = a_0 - \frac{a_1}{2} + \frac{a_2}{3} = A;$$

$$\int_{0}^{1} f(x)\mathrm{d}x = \left[a_0 x + \frac{a_1}{2} x^2 + \frac{a_2}{3} x^3\right]_{0}^{1} = a_0 + \frac{a_1}{2} + \frac{a_2}{3} = B;$$

$$\int_{1}^{2} f(x)\mathrm{d}x = \left[a_0 x + \frac{a_1}{2} x^2 + \frac{a_2}{3} x^3\right]_{1}^{2} = a_0 + \frac{3}{2} a_1 + \frac{7}{3} a_2 = C.$$

Rezultă următorul sistem de trei ecuații:

$$\left.\begin{aligned} a_0 - \frac{1}{2} a_1 + \frac{1}{3} a_2 &= A \\ a_0 + \frac{1}{2} a_1 + \frac{1}{3} a_2 &= B \\ a_0 + \frac{3}{2} a_1 + \frac{7}{3} a_2 &= C \end{aligned}\right\}$$

Scăzînd din fiecare ecuație pe cea precedentă, obținem

$$a_1 = B - A = \Delta F_{-1},$$

$$a_1 + 2a_2 = C - B = \Delta F_0,$$

în care, $\triangle F$ reprezintă sporul frecvenței relative.

Din nou scădem prima ecuație din a doua și obținem

$$2a_2 = (C - B) - (B - A) = \Delta^2 F$$

de unde

$$a_2 = \frac{1}{2} \Delta^2 F.$$

Să luăm acum semisuma primelor două ecuații inițiale:

$$a_0 + \frac{1}{3}\, a_2 = \frac{1}{2}(A + B),$$

$$a_0 = \frac{1}{2}(A+B) - \frac{1}{3}\, a_2 = \frac{A + B}{2} - \frac{\Delta^2 F}{6}.$$

Prin urmare,

$$f(x) = \frac{A + B}{2} - \frac{\Delta^2 F}{6} + \Delta F_{-1} \cdot X + \frac{\Delta^2 F}{2} \cdot x^2. \qquad \text{(IX,6.1)}$$

În limita părții de interval ce se descompune, de la 0 la $\theta < 1$, avem frecvența:

$$F_0^{\theta} = \int_0^{\theta} f(x)\, dx = \left[a_0 x + \frac{a_1 x^2}{2} + \frac{a_2 x^3}{3} \right]_0^{\theta} = a_0\, \theta + \frac{a_1}{2}\, \theta^2 + \frac{a_2}{3}\, \theta^3 =$$

$$= \theta \left(a_0 + \frac{a_1}{2}\, \theta + \frac{a_2}{3}\, \theta^2 \right) = \theta \left[\frac{A + B + \Delta F_{-1}\theta}{2} - \frac{\Delta^2 F(1 - \theta^2)}{6} \right]. \qquad \text{(IX,6.2)}$$

Această formulă a fost folosită pentru determinarea frecvenței în exemplul de mai sus.

Metoda folosită de Boiarski se încadrează într-o metodă mai generală.

Fie $f(x, \alpha_1, \alpha_2 \ldots, \alpha_m)$ o funcție de x care depinde de m parametri.

Admitem că curba

$$y = f(x;\ \alpha_1,\ \alpha_2,\ \ldots, \alpha_m)$$

este distribuția teoretică.

Întrucît avem m parametri, pentru determinarea lor numerică avem nevoie de m condiții, care se obțin în modul următor: trasăm o serie de drepte paralele cu axa Oy astfel ca porțiunile determinate de Ox, paralele cu Oy și respectiv curba căutată și curba empirică dată să fie respectiv egale. Se consideră atîtea arii parțiale cîte constante arbitrare intră în funcția $f(x; \alpha_1, \alpha_2, \ldots, \alpha_m)$, adică în cazul nostru m.

În exemplul tratat mai sus a fost considerată funcția

$$f(x) = a_0 + a_1 x + a_2 x^2,$$

conținînd trei parametri a_0, a_1, a_2.

Vom aplica acum această metodă, considerînd funcția

$$f(x) = ax + b,$$

care conține doi parametri. Avînd doi parametri, trebuie să folosim două condiții.

Vom rezolva problema într-un caz general, particularizînd apoi datele pentru exemplul nostru.

Fie x_0, x_1, x_2 trei puncte date pe axa Ox și y_1, y_2 frecvențele corespunzătoare intervalelor (x_0, x_1), (x_1, x_2).

Frecvența curbei teoretice în intervalul (x_0, x_1) este

$$\int_{x_0}^{x_1} (ax + b)\, dx = \frac{a}{2} \left(x_1^2 - x_0^2 \right) + b(x_1 - x_0).$$

Egalînd frecvența aceasta cu frecvența distribuției date, avem

$$\frac{a}{2} \left(x_1^2 - x_0^2 \right) + b\,(x_1 - x_0) = y_1$$

Să luăm acum semisuma primelor două ecuații inițiale:

$$a_0 + \frac{1}{3}\, a_2 = \frac{1}{2}(A + B),$$

$$a_0 = \frac{1}{2}(A+B) - \frac{1}{3}\, a_2 = \frac{A+B}{2} - \frac{\Delta^2 F}{6}.$$

Prin urmare,

$$f(x) = \frac{A+B}{2} - \frac{\Delta^2 F}{6} + \Delta F_{-1} \cdot X + \frac{\Delta^2 F}{2} \cdot x^2. \qquad \text{(IX,6.1)}$$

În limita părții de interval ce se descompune, de la 0 la $\theta < 1$, avem frecvența:

$$F_0^{\theta} = \int_0^{\theta} f(x)\, dx = \left[a_0 x + \frac{a_1 x^2}{2} + \frac{a_2 x^3}{3} \right]_0^{\theta} = a_0\, \theta + \frac{a_1}{2}\, \theta^2 + \frac{a_2}{3}\, \theta^3 =$$

$$= \theta \left(a_0 + \frac{a_1}{2}\, \theta + \frac{a_2}{3}\, \theta^2 \right) = \theta \left[\frac{A + B + \Delta F_{-1}\theta}{2} - \frac{\Delta^2 F(1 - \theta^2)}{6} \right]. \qquad \text{(IX,6.2)}$$

Această formulă a fost folosită pentru determinarea frecvenței în exemplul de mai sus. Metoda folosită de Boiarski se încadrează într-o metodă mai generală.

Fie $f(x, \alpha_1, \alpha_2 \ldots, \alpha_m)$ o funcție de x care depinde de m parametri.

Admitem că curba

$$y = f(x;\ \alpha_1,\ \alpha_2,\ \ldots,\ \alpha_m)$$

este distribuția teoretică.

Întrucît avem m parametri, pentru determinarea lor numerică avem nevoie de m condiții, care se obțin în modul următor: trasăm o serie de drepte paralele cu axa Oy astfel ca porțiunile determinate de Ox, paralele cu Oy și respectiv curba căutată și curba empirică dată să fie respectiv egale. Se consideră atîtea arii parțiale cîte constante arbitrare intră în funcția $f(x; \alpha_1, \alpha_2, \ldots, \alpha_m)$, adică în cazul nostru m.

În exemplul tratat mai sus a fost considerată funcția

$$f(x) = a_0 + a_1 x + a_2 x^2,$$

conținînd trei parametri a_0, a_1, a_2.

Vom aplica acum această metodă, considerînd funcția

$$f(x) = ax + b,$$

care conține doi parametri. Avînd doi parametri, trebuie să folosim două condiții.

Vom rezolva problema într-un caz general, particularizînd apoi datele pentru exemplul nostru.

Fie x_0, x_1, x_2 trei puncte date pe axa Ox și y_1, y_2 frecvențele corespunzătoare intervalelor (x_0, x_1), (x_1, x_2).

Frecvența curbei teoretice în intervalul (x_0, x_1) este

$$\int_{x_0}^{x_1} (ax + b)\, dx = \frac{a}{2}\left(x_1^2 - x_0^2\right) + b(x_1 - x_0).$$

Egalînd frecvența aceasta cu frecvența distribuției date, avem

$$\frac{a}{2}\left(x_1^2 - x_0^2\right) + b\,(x_1 - x_0) = y_1$$

Scanned with OKEN Scanner

și în mod analog

$$\frac{a}{2}\left(x_2^2 - x_1^2\right) + b\,(x_2 - x_1) = y_2.$$

Pentru determinarea parametrilor a și b avem prin urmare de rezolvat sistemul

$$\left.\begin{aligned} \frac{a}{2}\left(x_1^2 - x_0^2\right) + b\,(x_1 - x_0) &= y_1, \\ \frac{a}{2}\left(x_2^2 - x_1^2\right) + b\,(x_2 - x_1) &= y_2. \end{aligned}\right\} \qquad \text{(IX,6.3)}$$

În cazul nostru, folosind aceleași notații ca în primul exemplu, putem alege

$$x_0 = -1,\ x_1 = 1,\ x_2 = 2,$$

$$y_1 = 34{,}88 + 25{,}58 = 60{,}46,\ y_2 = 25{,}58.$$

Rezolvînd sistemul (IX,6.3), obținem valorile $a = -10{,}11$ și $b = 30{,}23$

$$\int_0^{0{,}5} (-10{,}11\,x + 30{,}23)\,dx = 13{,}46.$$

Prin urmare, prin acest procedeu de calcul rezultă că în aproximativ 13,46 % din 25,58 % sovhozuri, volumul producției globale pe cap de lucrător a variat între 25,1 și 30 mii ruble.

În modul cel mai natural și cel mai simplu, această problemă se rezolvă folosind frecvența cumulată.

Notăm prin $F(x)$ frecvența acumulată pînă la x. Cu aceleași notații ca mai sus putem lua pentru exemplul nostru :

$$F(-1) = 0,\ F(0) = 34{,}88,\ F(1) = 60{,}46,\ F(2) = 75{,}58.$$

Se cere să determinăm valoarea lui $F\left(\frac{1}{2}\cdot\right)$ Ajungem astfel la o problemă clasică de interpolare căreia putem să-i aplicăm orice formulă de interpolare. De exemplu vom aplica formula de interpolare a lui Lagrange (vezi Cap. XIII, punctul 6).

Aplicînd această formulă, obținem

$$F(x) = 34{,}88\,\frac{(x+1)(x-1)(x-2)}{2} = 60{,}46\,\frac{x(x+1)(x-2)}{2} +$$

$$+ 75{,}58\,\frac{x(x+1)(x-1)}{6}.$$

Făcînd $x = \frac{1}{2}$, rezultă

$$F\left(\frac{1}{2}\right) = 48{,}90.$$

Frecvența cuprinsă între 0 și $\frac{1}{2}$ este deci egală cu

$$F\left(\frac{1}{2}\right) - F(0) = 14{,}02.$$

Folosind deci trei metode diferite de aproximare am ajuns la valorile 14,02%, 13,46%, 14,02%. Rezultatele sînt destul de apropiate, între frecvența maximă și frecvența minimă existînd o diferență de 0,56 %.

2°. În cazul unei distribuții cu intervale inegale, tehnica de calcul se complică din cauză că sistemul de ecuații pentru parametrii a_0, a_1 și a_2 este mai puțin comod de rezolvat.

Să presupunem că avem următoarele grupe de vîrstă ale unei populații, cu frecvențele respective:

12—14 ani cu frecvența relativă $A = 7{,}9\%$,
15—19 ani „ „ „ $B = 8{,}9\%$,
20—29 ani „ „ „ $C = 18{,}0\%$,

și se cere să determinăm, din cadrul grupei a doua, frecvența relativă a populației de vîrstă cuprinsă între 15 și 16 ani inclusiv. După cum vedem, în grupa a doua avem un interval de 5 ani, în grupa următoare de 10 ani, iar în cea precedentă de 3 ani.

Considerînd din nou originea de variație a caracteristicii în limita inferioară a intervalului al doilea (15 ani), vom avea pentru grupa cuprinsă între 15 și 16 ani (care reprezintă 0,2 din mărimea întregului interval de 5 ani), următoarea frecvență relativă:

$$F_0^{0.2} = 0{,}2.\,(11{,}117 - \frac{5{,}933}{2} - 0{,}2 + \frac{2{,}25}{3} \cdot 0{,}04) \cong 2{,}11\%.$$

Demonstrația acestei formule de calcul se poate face în felul următor:
Pentru prima grupă avem

$$\int_{-0{,}6}^{0} f(x)\,dx = \left[a_0 x + \frac{a_1}{2} x^2 + \frac{a_2}{3} x^3\right]_{-0{,}6}^{0} = 0{,}6\,a_0 - 0{,}18\,a_1 + 0{,}072\,a_2 = A;$$

pentru grupa a doua (pe care o descompunem)

$$\int_0^1 f(x)\,dx = a_0 + \frac{1}{2} a_1 + \frac{1}{3} a_2 = a_0 + 0{,}5\,a_1 + 0{,}333\,a_2 = B;$$

pentru grupa a treia

$$\int_1^3 f(x)\,dx = \left[a_0 x + \frac{a_1}{2} x^2 + \frac{a_2}{3} x^3\right]_1^3 = 2\,a_0 + 4\,a_1 + \frac{26}{3} a_2 = C.$$

Prin urmare, vom avea sistemul de ecuații

$$\begin{aligned} 0{,}6\,a_0 - 0{,}18\,a_1 + 0{,}072\,a_2 &= 7{,}9, \\ a_0 + 0{,}5\,a_1 + 0{,}333\,a_2 &= 8{,}9, \\ 2\,a_0 + 4\,a_1 + 8{,}667\,a_2 &= 18{,}0, \end{aligned}$$

care ne dă

$$a_0 = 11{,}117,\ a_1 = -5{,}933 \text{ și } a_2 = 2{,}25$$

și atunci

$$F_0^{0.2} = \int_0^{0.2} f(x)\,dx = 0{,}2\left(a_0 + \frac{a_1}{2} 0{,}2 + \frac{a_2}{3} 0{,}2^2\right).$$

Așadar, din 8,9% — frecvența relativă a grupei de vîrstă de 15—19 ani — trebuie să considerăm că 2,11% reprezintă o populație în vîrstă de 16 ani, iar 6,79% peste 16 ani.

7. Reprezentarea grafică a distribuțiilor cu o singură caracteristică

În studiul distribuțiilor cu o singură caracteristică reprezentarea grafică este un auxiliar de primă importanță. Reprezentarea grafică o putem considera chiar ca o metodă specifică statisticii cu ajutorul căreia se pot observa mai ușor unele proprietăți ale dependenței funcționale dintre variabila independentă și frecvențele distribuției. Utilitatea ei se explică prin faptul că înlesnește analiza rapidă a fenomenelor cercetate, organul de simț al ochiului fiind în măsură a face rapid o apreciere a formei, direcțiilor și lungimilor, și mai ales, a raportului dintre acestea din urmă. Este evident că, caracterul manifestării unui fenomen de masă se poate desprinde și din șirul de date numerice cuprinse într-o serie statistică, aceasta se obține însă cu un efort de atenție mai mare, iar în unele cazuri devine chiar o imposibilitate.

Compararea manifestărilor unui fenomen în condiții diferite — care, după cum știm, este metoda fundamentală în cercetările statistice — este mult ușurată de reprezentarea grafică; ea face posibilă cuprinderea simultană a particularităților celor două serii numerice comparate.

Cu alte cuvinte, reprezentarea grafică, elaborată desigur cu respectarea regulilor cu caracter general stabilite de statistica teoretică, constituie un instrument util de cercetare.

Reprezentarea grafică este utilă și din alte două puncte de vedere. În primul rînd, ea poate fi folosită pentru găsirea în mod grafic a unor valori care ar necesita operațiuni complicate și uneori greoaie, dacă ar fi calculate direct (este vorba, evident, de unele valori aproximative, cînd nu este necesar un mare grad de precizie). În al doilea rînd, ea poate fi folosită și ca mijloc de control al îndeplinirii unor procese planificate, dat fiind că orice rezultat anormal este scos în evidență de graficul întocmit și prin aceasta impune o verificare și luarea măsurilor corespunzătoare de îndreptare. Cu acest prilej se pot corecta, eventual, erorile strecurate la culegerea datelor sau — în lipsa acestora — se pot depista cauzele anomaliei constatate.

A. *Reprezentarea grafică a distribuțiilor în spațiu.* Forma cea mai obișnuită de reprezentare grafică a distribuțiilor în spațiu este așa-numita *cartodiagramă.* Ea se face cu ajutorul unei hărți geografice — de aici și termenul de cartodiagramă — pe care sînt reprezentate datele statistice, prin suprafețe proporționale cu ele. Distribuția în spațiu corespunzătoare este reprezentată astfel prin imagini grafice ale acestor date, răspîndite pe hartă. Cel mai utilizat procedeu este acela al imaginilor formate din dreptunghiuri sau cercuri (se folosesc și alte figuri geometrice sau simboluri naturale).

Printr-un grafic cu folosirea cercurilor se poate reprezenta, de pildă, importul unei țări în raport cu țara de origine. Prin semicercuri de o

anumită culoare (sau marcate printr-un anumit semn) se reprezintă cantitatea importată din țara corespunzătoare, raportată la cantitatea totală a importului, iar prin semicercuri de altă culoare (sau marcat printr-un alt semn convențional) se reprezintă aceiași indicatori în expresie bănească.

Dintr-o singură privire asupra unei asemenea cartodiagrame se degajează semnificația datelor reprezentate.

B. *Reprezentarea grafică a distribuțiilor de o caracteristică calitativă.* Reprezentarea acestor distribuții constituie un capitol deosebit de important al reprezentării grafice, dat fiind că ilustrează, prin desene, anumite rapoarte numerice: Graficul corespunzător poartă numele de *diagramă.*

Să considerăm, de pildă, distribuția populației active din ramura industriei R.P.R., pe grupe sociale (tabelul 17).

Tabelul 17

Distribuția populației active din ramura industriei R.P.R. pe grupe sociale (după datele recensământului populației din 21 februarie 1956)

Nr. de ordine	Grupa socială	Frecvența absolută (F)	Frecvența relativă (%) (f)
A	B	1	2
	Total industrie	1 478 025	100,0
1	Muncitori	970 275	65,6
2	Funcționari *)	217 576	14,7
3	Meșteșugari cooperativizați	74 900	5,1
4	Colectiviști	2 906	0,2
5	Întovărășiți	1 787	0,1
6	Țărani individuali	57 625	3,9
7	Meșteșugari individuali **)	152 787	10,3
8	Liber profesioniști	40	+***)
9	Comercianți	14	+
10	Alții	64	+
11	Nedeclarată	51	+

*) Inclusiv ingineri, tehnicieni și personal de specialitate.

**) Inclusiv patronii unităților care au salariați.

***) Sub 0,1%.

Sursa : *Recensământul populației din 21 februarie 1956*, vol. 2, Cap. LVIII.

Diagrama prin care s-ar putea ilustra datele din acest tabel poate să ia formele cele mai variate :

1°. Numărul populației active din industrie pe fiecare grupă socială poate fi reprezentat prin diferite figuri geometrice, ca pătrate, cercuri, cuburi, ale căror suprafețe, respectiv volume, să fie proporționale — pe grafic — cu frecvențele corespunzătoare.

Pentru asigurarea acestei proporționalități, trebuie extrase rădăcinile pătrate ale frecvențelor (pentru a construi laturile pătratelor sau razele cercurilor care vor fi proporționale cu aceste rădăcini), sau rădăcina cubică a frecvențelor (în cazul reprezentării prin cuburi).

Scanned with OKEN Scanner

În acest mod de reprezentare grafică, raportul numeric dintre frecvențe, care reprezintă cel mai important avantaj al reprezentării grafice, nu este așa de ușor perceptibil ca în cazul raportului de lungime. El este totuși adeseori utilizat mai ales atunci cînd frecvențele diferitelor grupe trebuie repartizate în diferite puncte cum ar fi cazul unei distribuții în spațiu (vezi reprezentarea din fig. 11, prin pătrate).

Notă: Urmează reprezentarea prin pătrate a datelor din tabelul 11, ultimele patru grupe fiind reunite în una singură cu denumirea generică de „alții"; s-au reunit de asemenea grupele 4 și 5 în total fiind șapte grupe.

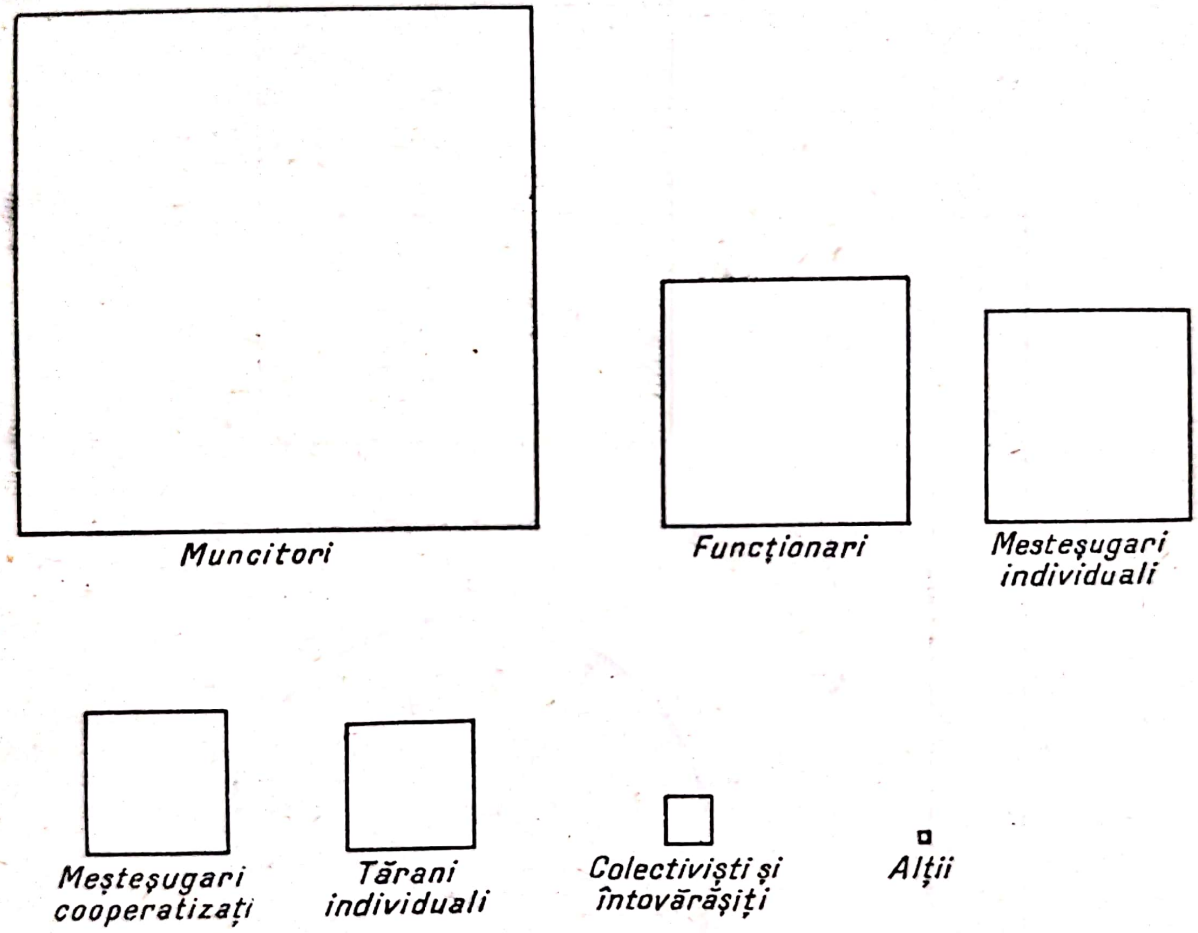


Fig. 11

2°. Frecvențele distribuției din tabelul 17 pot fi reprezentate printr-un dreptunghi împărțit în mai multe alte dreptunghiuri divizionare. În acest caz înălțimea dreptunghiurilor divizionare variază proporțional cu mărimea frecvențelor respective. Acest tip de grafic are avantajul de a înlătura inconvenientele celui precedent, fiind vorba de a fi apreciată numai o singură dimensiune.

Impresiile pe care ni le lasă diviziunile A și B din fig. 12 sînt diferite. Diagrama A (cu baza dreptunghiului îngustă) ne lasă impresia că frecvențele mici sînt destul de importante, în schimb, diagrama B exagerează importanța numerică a frecvențelor mari în dauna celor mici.

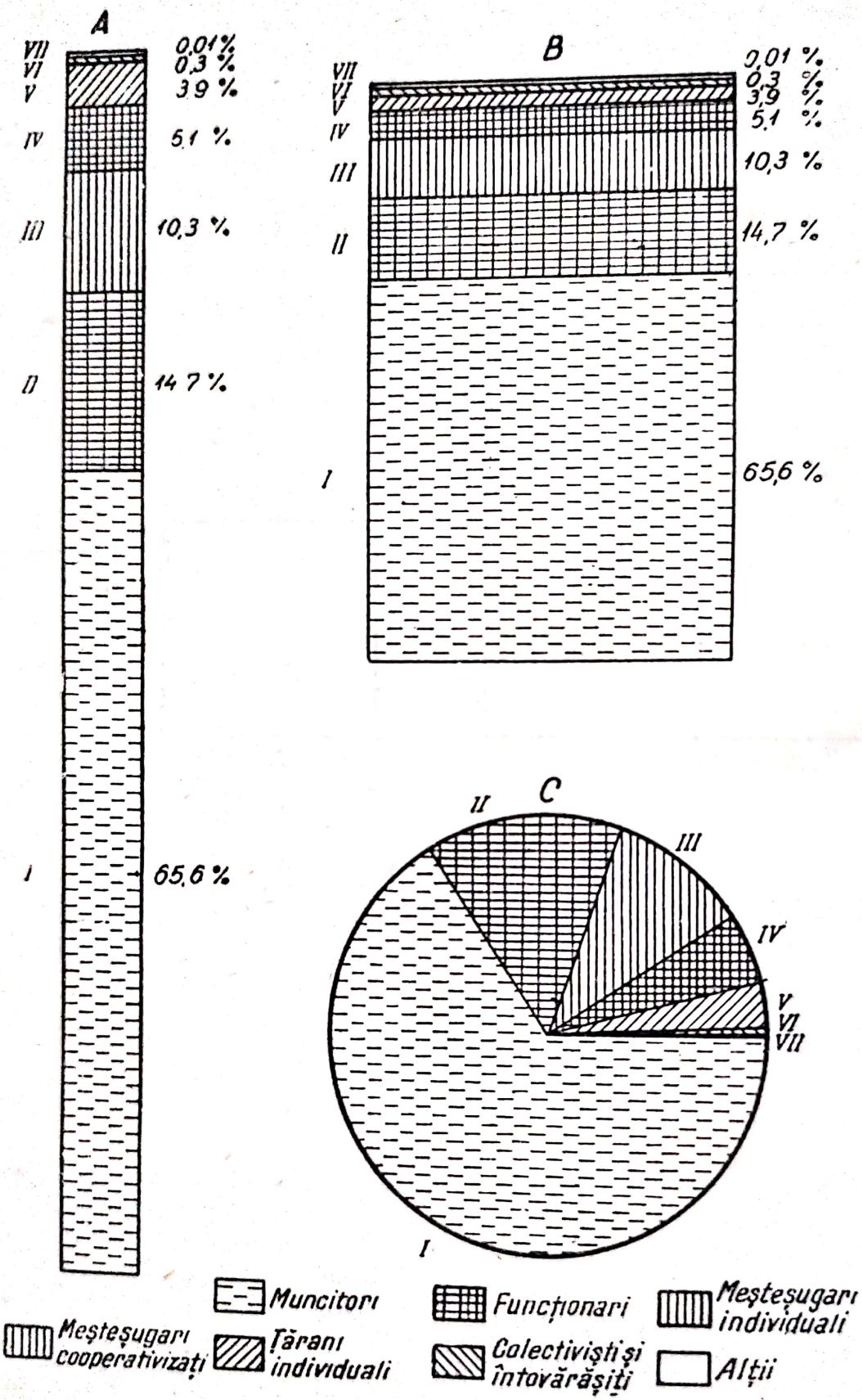


Fig. 12

Scanned with OKEN Scanner

De aceeași categorie este și diagrama în cerc în care frecvențele sînt reprezentate prin sectoare de cerc (numărul de grade al arcelor corespunzătoare este proporțional cu mărimea frecvențelor). Diagramele în cerc sau dreptunghiulare se mai numesc și *diagrame de structură*.

3°. Datele din tabelul 17 mai pot fi reprezentate prin niște dreptunghiuri de aceeași bază și de înălțimi proporționale cu frecvențele. Dreptun-

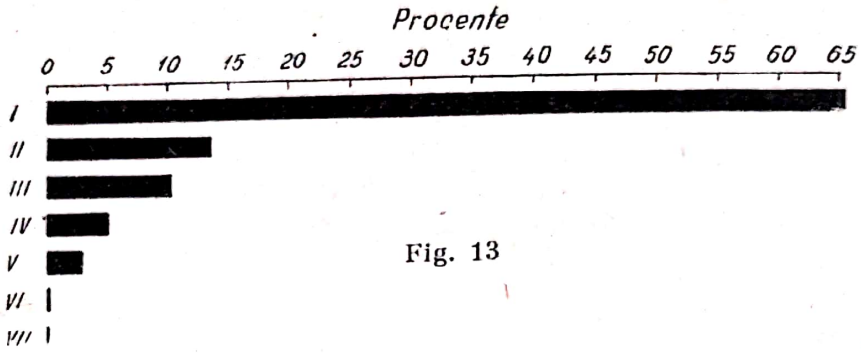


Fig. 13

ghiurile sînt așezate vertical sau orizontal (ca în fig. 13), avînd alăturată o scară care pune în evidență raporturile dintre frecvențe.

Menționăm că așezarea verticală a dreptunghiurilor este preferabilă în cazurile cînd între diferitele grupe ale distribuției se poate stabili o ordine graduală.

După cum se vede, principiul reprezentării unei distribuții prin diagrame este pe cît de simplu pe atît de sugestiv. Tocmai aceasta este și regula generală care trebuie respectată la construirea diagramelor: simplitate și claritate.

C. *Reprezentarea grafică a seriilor de variație.* Pentru reprezentarea acestor serii ne folosim tot de sistemul de referență al axelor rectangulare, introduse de Descartes, pe care l-am folosit la reprezentarea grafică a distribuțiilor de caracteristici calitative și-l vom mai folosi la reprezentarea seriilor în timp. Valorile caracteristicii independente în raport cu care s-a făcut distribuția se notează pe axa absciselor, iar în dreptul lor, în ordonată, se reprezintă valoarea frecvențelor corespunzătoare.

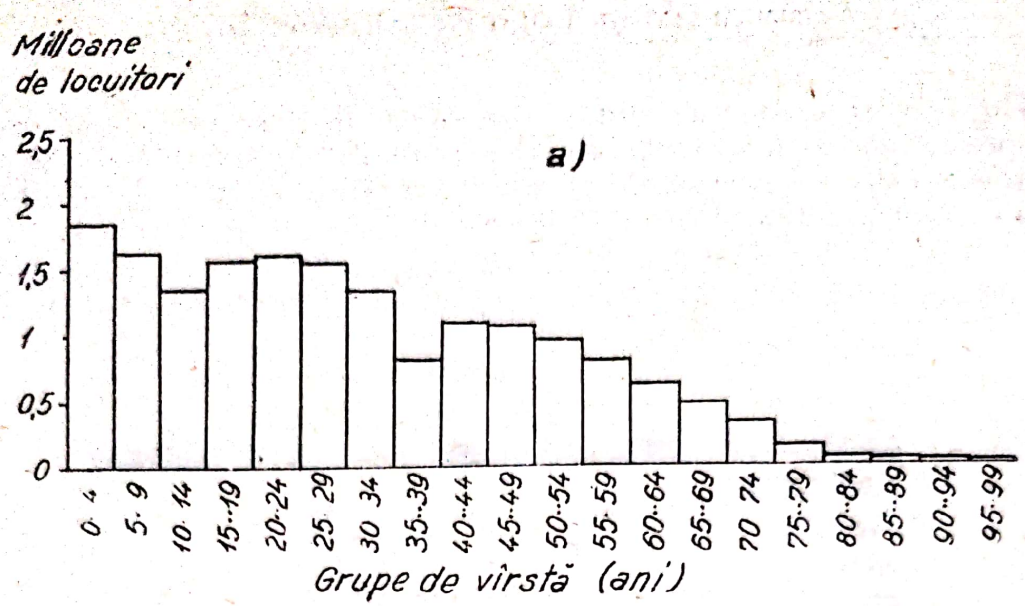
Există trei moduri de reprezentare a seriilor de variație care se mențin — cu mici deosebiri de formă — în cele două ipoteze posibile: serii cu intervale egale și serii cu intervale inegale.

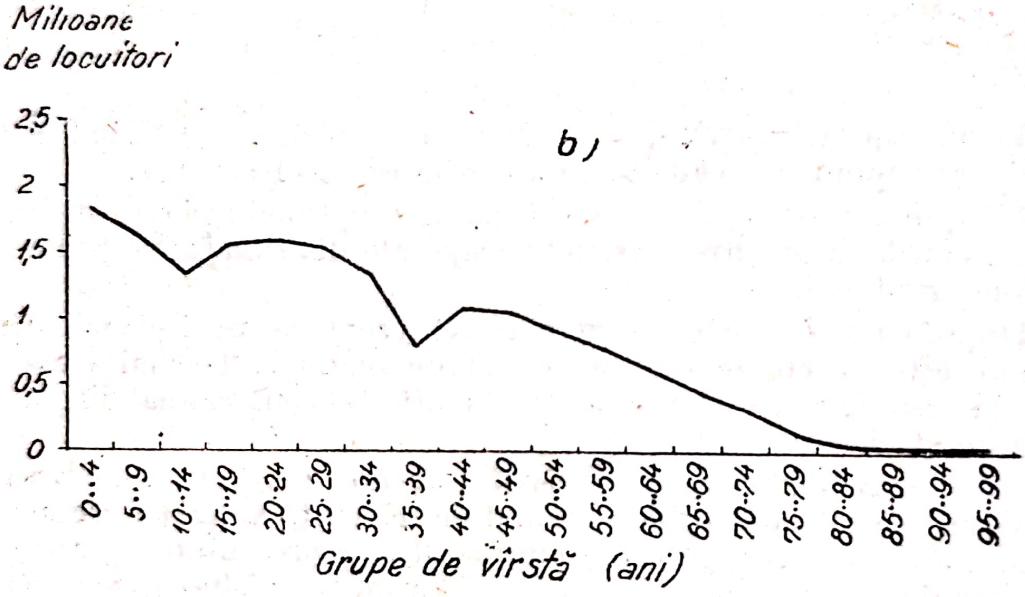
1°. În ipoteza că avem o serie cu *intervale egale* ca cea ilustrată în tabelul 13, o primă modalitate de a o reprezenta grafic constă în a construi pe fiecare din intervalele de grupare ale caracteristicii cantitative, dreptunghiuri de înălțimi proporționale cu frecvențele distribuției.

Un astfel de grafic se numește *histogramă* *).

În figura 14 a) este arătată distribuția pe grupe de vîrstă a populației țării noastre, pe baza datelor din tabelul 13.

*) Acest termen a fost introdus de Karl Pearson.





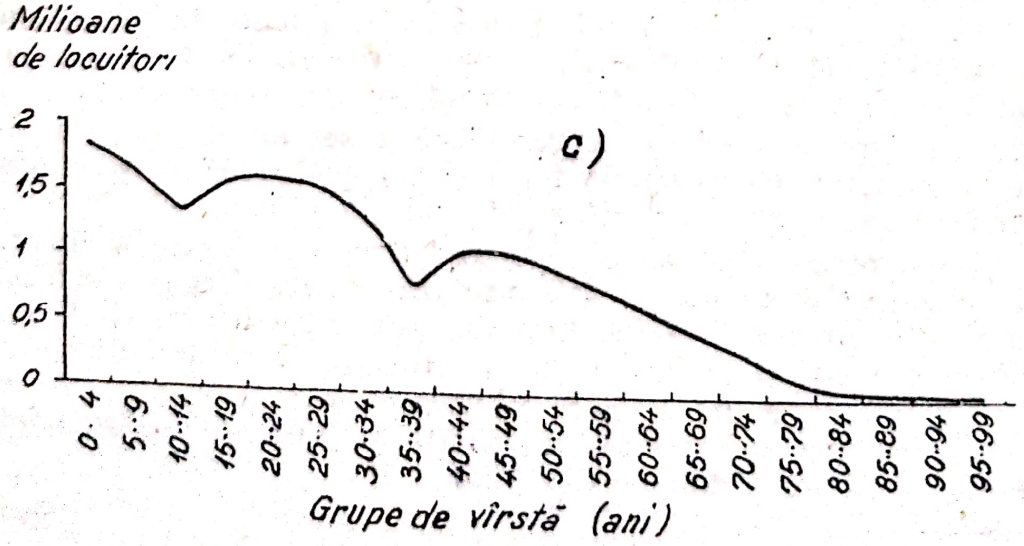


Fig. 14

Scanned with OKEN Scanner

După cum se vede, histograma este reprezentarea grafică rezultată din considerarea frecvențelor distribuției ca arii ale unor dreptunghiuri. De aici rezultă că histograma se întrebuințează pentru reprezentarea grafică a distribuțiilor avînd valorile caracteristicii grupate pe intervale proporționale cu baza dreptunghiurilor.

Un alt mod de reprezentare este *poligonul frecvențelor*. Acest grafic se construiește în felul următor: pe mijlocul intervalelor de grupare ale caracteristicii variabile, reprezentată pe axa absciselor, se ridică perpendiculare proporționale cu frecvențele și apoi se unesc extremitățile acestor perpendiculare printr-o linie poligonală (vezi fig. 14b)).

Prin urmare, linia poligonală unește punctele ale căror abscise și ordonate sînt, respectiv, — perechile de valori luate de caracteristica studiată și frecvența corespunzătoare.

Din cele arătate mai sus se observă că se poate trece ușor de la histograma distribuției la poligonul frecvențelor și invers. Într-adevăr, unind prin segmente de dreaptă mijlocul bazelor superioare ale dreptunghiurilor histogramei, se obține poligonul frecvențelor. Printr-o operație inversă se poate construi ușor histograma distribuției.

În sfîrșit, prin punctele determinate pentru a fi trasă linia poligonală se poate figura și o *curbă* trecînd prin aceste puncte sau aproape de ele — numită *curba frecvențelor* sau *curba de distribuție* a cărei semnificație este de o importanță deosebită în statistică (vezi fig. 14c)).

Curba de distribuție reprezintă tocmai acea distribuție teoretică continuă care după cum am văzut în domeniul fenomenelor și proceselor stohastice, este dată de *legea de probabilitate* respectivă.

Determinarea curbei frecvențelor pornind de la datele statistice primare — cu toate consecințele fructoase ce decurg de aici pentru formularea unor concluzii juste cu privire la trăsăturile ei caracteristice, valorile ei tipice etc. — constituie una din cele mai importante și totodată cele mai interesante probleme de statistică.

Trebuie să menționăm că poligonul și curba de frecvențe fac posibilă, în primul rînd, precizarea *formei* pe care o are distribuția studiată; în al doilea rînd, ele ușurează compararea grafică a două serii paralele, relevînd, eventual, dacă există o interdependență *reală* între fenomenele oglindite în aceste serii și corelația dintre ele. Funcția cognitivă a histogramei este oarecum diferită: în comparație cu graficele precedente ea relevă mai pregnant diferența de mărime între frecvențele distribuției, aceasta, însă, în dauna unei reprezentări mai sugestive a formei ei generale.

2°. În ipoteza că avem o serie cu *intervale inegale*, reprezentarea ei grafică necesită o transformare specială.

Să considerăm, de pildă, distribuția localităților urbane din R.P.R. după numărul de locuitori la 1 iulie 1959 (vezi tabelul 18), pe care vrem să o reprezentăm grafic printr-o histogramă.

În primul rînd, vom admite convențional *) că primul interval al seriei este închis, avînd limita inferioară egală cu 500 locuitori, iar ultimul,

*) Admitem această convenționalitate în scop strict ilustrativ, acela de a face mai clară metoda de reprezentare grafică a seriilor de variație cu intervale inegale.

Scanned with OKEN Scanner

de asemenea, avînd limita superioară egală cu 300 000 locuitori (excluzînd orașul București).

Întrunind apoi primele trei intervale într-unul singur*), vom observa că în noua distribuție primul interval va fi de 9 500 de locuitori, al doilea de 15 000, al treilea de 25 000 locuitori și tot așa mai departe, intervalele cresc, ultimul avînd 250 000 locuitori (aproximativ de 25 de ori mai mare decît primul).

Pentru a construi histograma acestei distribuții vom ține seama că pe acest tip de grafic frecvențele sînt reprezentate prin arii proporționale cu ele. Aceasta înseamnă, de pildă, pentru a face comparabil dreptunghiul reprezentînd frecvența intervalului 10 000—24 999 cu dreptunghiul corespunzător primului interval, avînd baza mai mică, de aproximativ 1,5 ori, că trebuie ca înălțimea dreptunghiului corespunzător intervalului al doilea să fie micșorată de aproximativ 1,5 ori, asigurîndu-se astfel proporționalitatea necesară. Se micșorează în mod corespunzător și frecvențele celorlalte intervale care, în comparație cu frecvențele primului interval, sînt de un anumit număr de ori mai mari.

Vom obține, astfel, o nouă distribuție (tabelul 18) cu frecvențele reduse la intervalul de 9 500 locuitori **).

Tabelul 18

Distribuția localităților urbane din R.P.R. după numărul de locuitori la 1 iulie 1959

Numărul locuitorilor	Frecvențele inițiale ale intervalelor	Frecvențe reduse la intervalul de 10 000 locuitori
0	1	2
500— 9 999	242	242
10 000— 24 999	75	50
25 000— 49 999	20	8
50 000— 99 999	5	1
100 000—350 000	11	0,44

Histograma construită pe baza frecvențelor reduse este ilustrată în figura 15.

Folosirea poligonului de frecvență pentru a reprezenta serii de variație cu intervale inegale nu este de recomandat. Se observă din figura 15 că, unind mijloacele bazelor superioare ale dreptunghiurilor reprezentînd frecvențele reduse, putem obține și un poligon de frecvențe, însă, la ultimile două grupe linia poligonală se confundă cu histograma.

Trebuie să precizăm acum că în cazul unei distribuții discrete, reprezentarea ei grafică nu se poate face decît prin histogramă, întrucît celelalte două tipuri de grafice — poligonul sau curba de frecvențe — ne-ar da impresia continuității caracteristicii considerate.

*) Întrunirea acestor grupe este necesară pentru a găsi unități de lungime potrivite la reprezentarea grafică a seriei care are o amplitudine a variației prea mare.

**) În reprezentarea grafică, pentru ușurarea calculelor vom considera frecvențele reduse la un interval egal cu 10 000 locuitori.

Scanned with OKEN Scanner

8. Forma distribuțiilor de variație

Vom prezenta acum succint cinci tipuri de forme ale unei serii de variație cu o singură caracteristică, apreciindu-le ca cele mai caracteristice din mulțimea de forme care se întîlnesc în practică. Caracterul fiecărei forme îl vom scoate în evidență aplicînd metoda grafică, ținînd

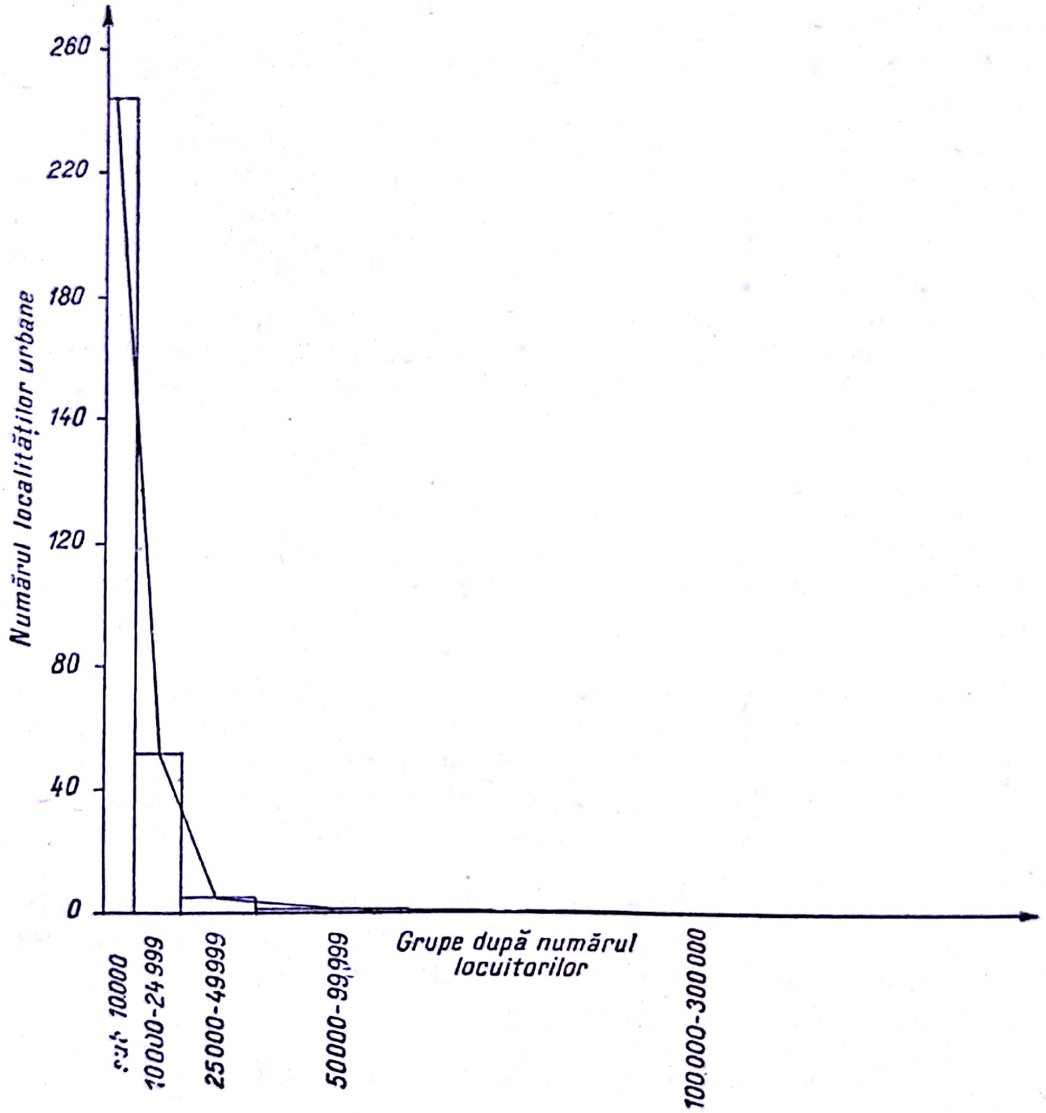


Fig. 15

seama de poziția pe care o ocupă pe scara de variație, intervalul căruia îi corespunde frecvența maximă.

A. *Distribuții simetrice.* Acestea sînt distribuții la care avem aceleași valori pentru frecvențele situate la distanțe egale de o parte și de alta a valorii de frecvență maximă. De fapt, nu există distribuții care să corespundă riguros acestei definiții, deoarece, în general, valorile frecvențelor distribuite în modul arătat față de frecvența maximă diferă, într-o oarecare măsură, unele de altele.

Reprezentate prin poligonul de frecvență corespunzător, distribuțiile simetrice se prezintă sub forma profilului unui clopot. Ținînd seama de mulțimea de cazuri cînd distribuțiile se prezintă sub una din formele diferite de cea a profilului unui clopot, putem considera că distribuțiile simetrice reprezintă un caz rar în statistică. Cele mai răspîndite distribuții simetrice sînt cunoscute sub numele de *distribuții normale.*

Pe plan teoretic, probabilistic, unei distribuții statistice normale îi corespunde cunoscuta lege normală de probabilitate pe care am examinat-o la capitolul VII și care, după cum știm, a fost studiată prima oară de Laplace și apoi de Gauss.

Exemple de distribuții simetrice ni le oferă, îndeosebi, antropometria, biometria, precum și unele măsurători și observații ale fenomenelor din natură — mai rar ale fenomenelor din viața social-economică.

Exemplul clasic de distribuție normală ni-l oferă distribuția înălțimilor dintr-o colectivitate omogenă.

Forma unei asemenea distribuții am văzut-o și în capitolul VII în care am considerat probabilitatea alegerii la întîmplare a unei valori x ce se supune legii normale; ea este dată sub forma funcției

$$g = \frac{1}{\sqrt{2\pi}.\sigma} e^{-\frac{(x-\bar{x})^2}{2\sigma^2}}$$

care reprezentată grafic are forma unei curbe simetrice.

În figura 16 reprezentăm distribuția înălțimii unei populații omogene (în cm) *).

Mai dăm următorul exemplu:

În vara anului 1935, în două regiuni din U.R.S.S. (Dnepropetrovsk și Saratov) s-a efectuat o cercetare statistică a greutății boabelor de grîu, culese de pe fiecare metru pătrat a ariilor de 1/4 ha. Ariile alese au fost omogene din toate punctele de vedere, pe fiecare din ele grîul a fost semănat și recoltat în condiții naturale și agrotehnice identice. Rezultatele măsurătorilor sînt prezentate în tabelul 19.

Distribuțiile din acest tabel aproape că se confundă cu cele două curbe teoretice corespunzătoare — curbe normale. Reprezentarea lor concretă este determinată de următoarele valori, tipice:

1°. Pentru regiunea Dnepropetrovsk:

$$\bar{x} = 80{,}67\,; \quad \sigma^2 = 325{,}03\,; \quad \sigma = 18{,}03.$$

2°. Pentru regiunea Saratov:

$$\bar{x} = 83{,}67\,; \quad \sigma^2 = 428{,}14\,; \quad \sigma = 20{,}69.$$

Apropierea dintre distribuția statistică și curba normală a lui Laplace-Gauss este ilustrată grafic în figura 18 în care datele caracterizînd situația din regiunea Dnepropetrovsk sînt reprezentate sub forma histogramei.

*) Acest exemplu de distribuție și următorul sînt luate după A.I. Boiarski, L.S. Brand și alții din lucrarea *Statistika*. Moskva, 1936. În fig. 16 este reprezentată distribuția după înălțime a tinerilor din U.R.S.S. — contingentul 1924.

Scanned with OKEN Scanner

Tabelul 19

Distribuția greutății boabelor de grîu din două regiuni din U.R.S.S. — în 1935

Grupe de variație (în grame)	Frecvența în %		Grupe de variație (în grame)	Frecvența în %	
	Reg. Dnepropetrovsk	Reg. Saratov		Reg. Dnepropetrovsk	Reg. Saratov
0	1	2	3	4	5
0— 10	0,04	0,00	110—120	2,76	4,70
10— 20	0,04	0,00	120—130	1,08	2,23
20— 30	0,08	0,05	130—140	0,48	1,39
30— 40	0,60	0,05	140—150	0,08	0,69
40— 50	3,08	2,57	150—160	0,00	0,25
50— 60	8,72	7,33	160—170	0,00	0,15
60— 70	15,12	15,50	170—180	0,04	0,00
70— 80	20,76	20,79	180—190	0,00	0,05
80— 90	21,60	20,69	190—200	0,00	0,00
90—100	16,64	14,90	200—210	0,00	0,05
100—110	8,88	8,51	210—220	0,00	0,10
			Total	100,00	100,00

B. *Distribuții ușor asimetrice.* Acestea sînt acele distribuții în care valorile frecvențelor situate la distanțe egale de o parte și de alta a frecvenței maxime diferă puțin între ele. Aceste tipuri de distribuții sînt numeroase și se întîlnesc în toate domeniile.

Forma unor asemenea distribuții o întîlnim, de pildă, la fenomenele legate de evoluția vieții.

În figura 18 este ilustrată o asemenea distribuție. Este vorba de distribuția bărbaților sub 60 de ani din R.P.R. căsătoriți, pe grupe de vîrstă, după datele recensămîntului din 21 februarie 1956 (vezi și tabelul 20).

Tabelul 20

Distribuția bărbaților căsătoriți din R.P.R. pe grupe de vîrstă, după datele recensămîntului din 21 februarie 1956

Grupe de vîrstă x	Frecvențe F	Frecvențe relative %
0	1	2
Total R.P.R.	4 196 396	100,0
Sub 14 ani	2	—
14—19 „	27 177	0,64
20—24 „	227 426	5,41
25—29 „	606 202	14,44
30—39 „	917 115	21,85
40—49 „	1 112 816	26,41
50—54 „	429 223	10,22
55—59 „	342 298	8,15
60 ani și peste și vîrste nedeclarate	574 139	13,68

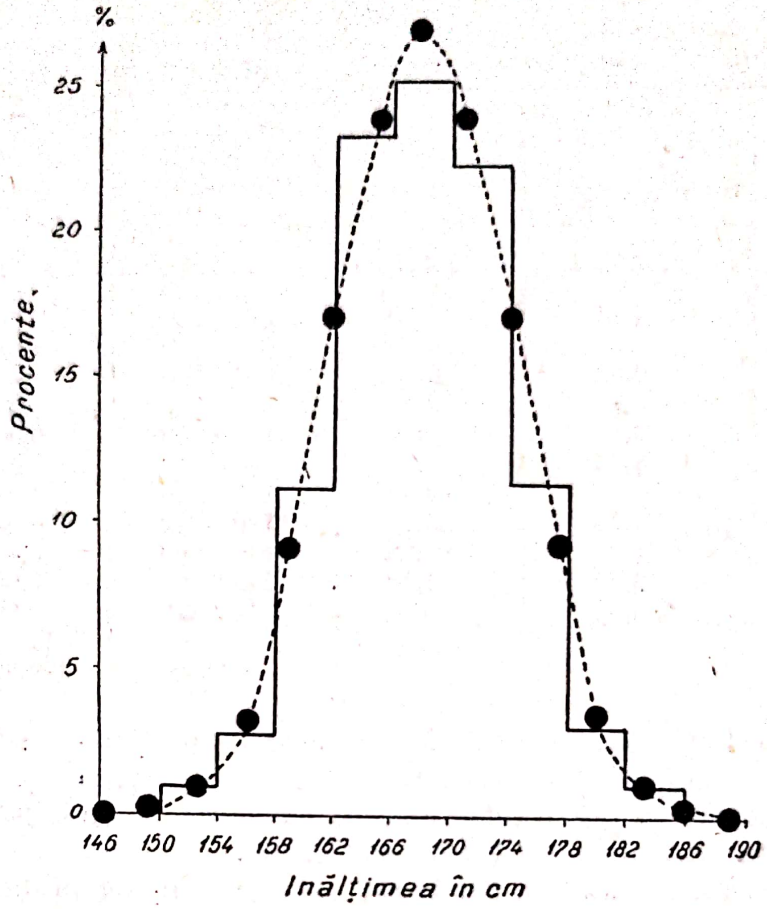

%
25
20
15
10
5
0
Procente.
146 150 154 158 162 166 170 174 178 182 186 190
Înălțimea în cm

Fig. 16

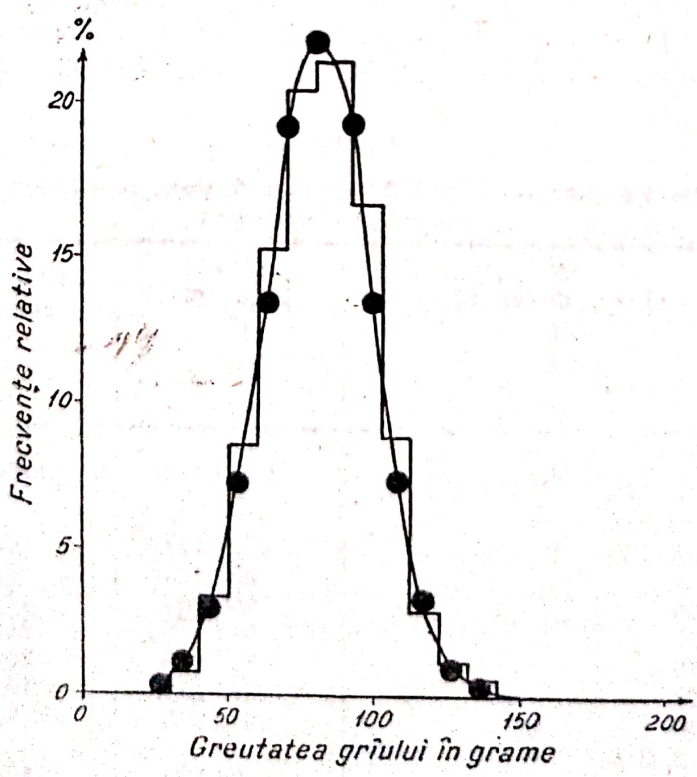

%
20
15
10
5
0
Frecvențe relative
0 50 100 150 200
Greutatea grîului în grame

Fig. 17

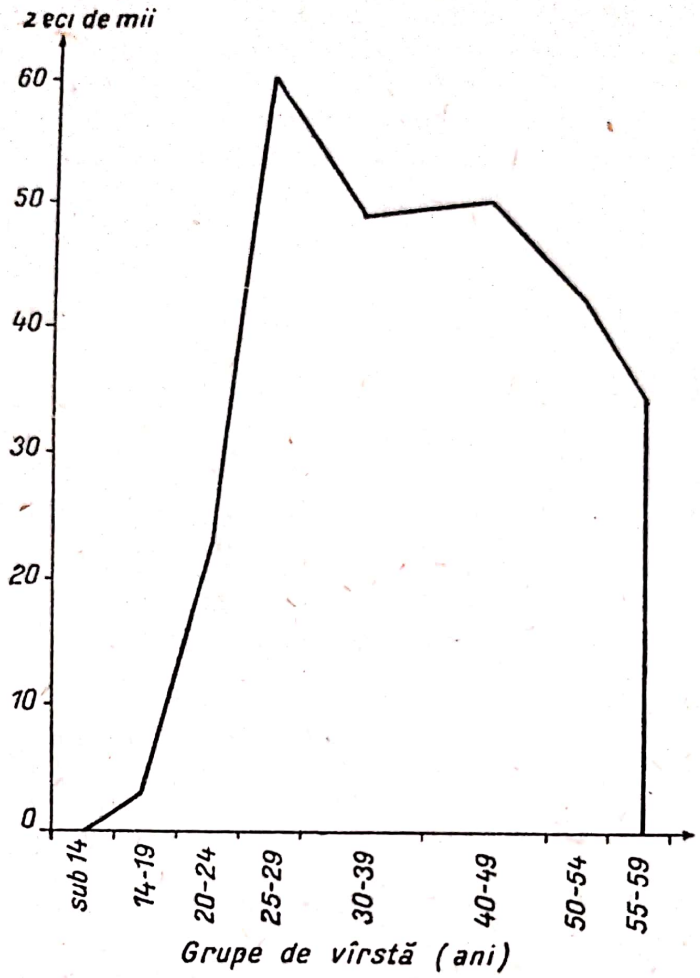


Fig. 18

C. *Distribuții pronunțat asimetrice.* Se prezintă foarte rar în statistică. Ele se caracterizează prin faptul că majoritatea frecvențelor sînt concentrate de o singură parte a frecvenței maxime care corespunde la o valoare a caracteristicii variabile foarte apropiată de unul din capetele intervalului de oscilație a distribuției.

Se poate întîmpla, uneori, ca din cauza unei grupări cu intervale prea mari, multe din distribuțiile din această categorie să fie de forma distribuțiilor *J*. În general, în cazul distribuțiilor pronunțat asimetrice se remarcă faptul că cele două extremități ale curbei frecvențelor nu sînt egale.

Distribuția S.M.T. după volumul mediu de lucrări pe un tractor, la sfîrșitul anului 1959 (vezi tabelul 9), are forma unei distribuții pronunțat asimetrice. O ilustrăm în graficul din figura 19.

D. *Distribuții în formă de J.* Sînt distribuții în care frecvențele sînt concentrate de o singură parte a frecvenței maxime, corespunzînd uneia din valorile caracteristicii din extremitățile scării de variație. Aceste distribuții trebuie cercetate cu multă atenție, deoarece unele distribuții

pronunțat asimetrice pot apărea ca distribuții în *J*. Astfel de situații se întîmplă, de obicei, în cazul distribuțiilor continue datorită faptului că împărțirea intervalului marginal în subintervale mai mici poate merge oricît de mult. Este evident că toate aceste considerații nu mai au obiect în cazul distribuțiilor discrete. Unii autori nu consideră distribuțiile în

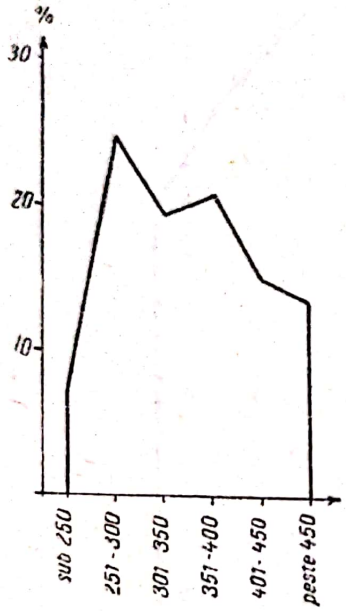


Fig. 19

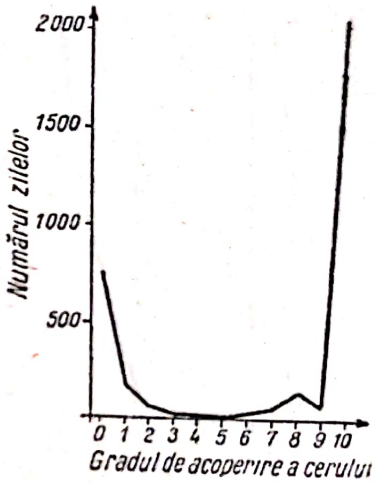


Fig. 20

J ca un tip de formă aparte, ci le înglobează în tipul de formă a distribuțiilor pronunțat asimetrice.

E. *Distribuții în U*. Aceste distribuții pot fi simetrice sau asimetrice (ușor sau mai pronunțat) în raport cu o anumită valoare corespunzînd celei mai mici frecvențe. În aceste distribuții, frecvențele referitoare la valorile (intervalele) marginale sînt maxime. Distribuțiile în *U* sînt dintre cele mai rare.

Forma unei distribuții în *U* este arătată în figura 20.

Scanned with OKEN Scanner

Capitolul X

Mărimi medii folosite în practica statisticii

1. Considerații generale

Din teoria generală a statisticii se știe că statistica dispune de diferite procedee cu ajutorul cărora prezintă datele obținute din observare, într-un mod adecvat scopului către care este îndreptată cercetarea.

Prezentarea datelor observate sub forma unui tabel statistic este primul procedeu prin care statisticianul desprinde din ansamblul datelor observate, caracteristicile fenomenului studiat. Un tabel statistic întocmit judicios exprimă necesitatea pentru fiecare cercetător de a reduce la un volum mic cantitatea considerabilă de date observate; numai în felul acesta el poate cuprinde esențialul din mulțimea datelor observate.

De asemenea, reprezentarea grafică a unei distribuții statistice care — după cum am văzut — scoate în evidență unele elemente specifice ale acesteia, este justificată de necesitatea de a compara între ele diferite distribuții, satisfăcîndu-se astfel exigențele de cunoaștere mai mari decît cele oferite de simpla înfățișare a datelor în tabele statistice. Utilizînd reprezentarea grafică se pot compara regiunile unei valori centrale care caracterizează concentrarea și dispunerea valorilor variabilei precum și gradul de asimetrie a distribuției.

Statisticii îi mai este proprie și o altă metodă și anume, aceea de a exprima sintetic, prin cîte o valoare numerică, un ansamblu de informații conținute în mulțimea de date observate. Asemenea valori sau mărimi medii se numesc *valori caracteristice ale distribuției* și cunoașterea lor poate determina în mod unic distribuțiile respective. Prin aceste valori (momente, funcții caracteristice, cumulanți etc.), statistica dispune de un instrument de cercetare și mai puternic decît acela pe care-l oferă tabelele statistice sau reprezentările grafice.

Folosind forma distribuției, intervalele de grupare etc., statisticianul va căuta să determine valorile ei caracteristice, caracterizînd astfel, prin

valori determinate unic, ordinul de mărime al datelor ce alcătuiesc o serie statistică.

Ordinul de mărime comun unui șir de valori poate fi determinat prin procedee diferite, conducînd, respectiv, la valori reprezentative de asemenea diferite, în funcție de proprietatea determinantă a distribuției pe care vrem s-o scoatem în evidență și s-o caracterizăm. În teoria și practica statisticii s-au considerat cîteva din aceste valori, cele mai reprezentative, denumite *valori tipice*.

Vom considera, în continuare, anumite proprietăți (funcții) *determinante* cu ajutorul cărora vom deduce diferite valori medii. Pentru ca o proprietate determinantă să fie oglindită prin valoarea medie respectivă este necesar ca șirul valorilor medii, care sînt valori abstracte, $\bar{x}, \bar{x}, \bar{x}, \ldots \bar{x}$, să nu se deosebească de șirul valorilor concrete, obținute din observare, $x_1, x_2, \ldots x_n$, în raport cu această proprietate.

Cu alte cuvinte, valoarea medie este o caracteristică abstractă a colectivității studiate care, dacă este pusă în locul valorilor individuale concrete, face ca funcția determinantă să rămînă neschimbată. Aceasta înseamnă că expresia cantitativă, numerică, a proprietății determinante pentru întreaga colectivitate rămîne aceeași și în cazul cînd în locul valorilor individuale este considerată valoarea lor medie.

Trebuie să precizăm, însă, că proprietatea determinantă și forma relației matematice care o exprimă sînt condiționate în întregime de legătura specifică lăuntrică, de esența *calitativă* a fenomenului studiat pe care le putem stabili — în domeniul fenomenelor sociale și economice — numai cu ajutorul tezelor materialismului istoric și ale economiei politice marxiste.

Ne vom opri acum, pe scurt, asupra valorilor medii tipice cele mai des folosite, ținînd seama că o serie de noțiuni și procedee tehnice de calcul matematic ne sînt de acum cunoscute din capitolele consacrate calculului probabilităților.

2. Media aritmetică și proprietățile ei

MEDIA ARITMETICĂ SIMPLĂ. Dacă considerăm funcția determinantă de forma

$$x_1 + x_2 + \ldots + x_n = n \cdot \bar{x},$$

atunci deducem imediat media aritmetică simplă:

$$\bar{x} = \frac{x_1 + x_2 + \ldots + x_n}{n} = \frac{\sum_{i=1}^{n} x_i}{n}. \qquad (X,2.1)$$

Această medie se folosește cînd variantele caracteristicii studiate se întîlnesc cîte o singură dată și au aceleași ponderi.

Așa, de exemplu, media aritmetică simplă a valorilor 8, 9, 11 și 12 este egală cu suma lor, împărțită la patru adică:

$$\frac{40}{4} = 10.$$

Cînd variantele caracteristicii au ponderi diferite, ceea ce înseamnă că se întîlnesc de mai multe ori în şirul de valori date, pentru calculul mediei lor se foloseşte media aritmetică ponderată pe baza aceleiaşi funcţii determinante.

MEDIA ARITMETICĂ PONDERATĂ. Dacă notăm cu $n_1, n_2, n_3, \ldots, n_n$ ponderile respective ale şirului de valori $x_1, x_2, x_3, \ldots, x_n$, vom avea funcţia determinantă de forma

$$x_1n_1 + x_2n_2 + \ldots + x_nn_n = \overline{x} \cdot \Sigma n_i$$

din care rezultă imediat formula mediei aritmetice ponderate

$$\overline{x} = \frac{\sum_{i=1}^{n} x_i n_i}{\sum_{i=1}^{n} n_i}. \qquad (X,2.2)$$

Dacă valorilor din exemplul de mai sus le atribuim anumite ponderi, să zicem, 3, 2, 1, 2, avem distribuţia :

x_i	n_i
$8 = x_1$	$3 = n_1$
$9 = x_2$	$2 = n_2$
$11 = x_3$	$1 = n_3$
$12 = x_4$	$2 = n_4$

avînd valoarea medie

$$\overline{x} = \frac{\sum_{i=1}^{n} x_i n_i}{\sum_{i=1}^{n} n_i} = \frac{8 \cdot 3 + 9 \cdot 2 + 11 \cdot 1 + 12 \cdot 2}{3 + 2 + 1 + 2} = \frac{77}{8} = 9{,}625.$$

Presupunînd că valorile x_i reprezintă salariul orar individual în lei, iar valorile n_i numărul respectiv al muncitorilor, fondul total de salarii ce se plăteşte acestui colectiv de muncitori — proprietatea determinantă pe care am luat-o în considerare — este egal cu

$$\overline{x} \cdot 8 = 9{,}625 \cdot 8 = 77 \text{ lei.}$$

Dacă însemnăm frecvenţele relative cu

$$\frac{n_i}{\Sigma n_i} = f_i,$$

formula (X,2.2) devine

$$\overline{x} = \Sigma f_i x_i, \qquad (X,2.3)$$

Scanned with OKEN Scanner

formă care are mai mult o importanță teoretică — deși mai simplă decît (X,2.2) — deoarece conduce pe un drum mai lung la valoarea lui $\bar{x}$, necesitînd calcularea prealabilă a frecvențelor relative.

PROPRIETĂȚILE MEDIEI ARITMETICE. Media aritmetică se bucură de o serie de proprietăți care se folosesc adeseori în practică pentru simplificarea calculului ei.

1°. Media unui șir de valori egale este egală cu acea valoare. Dacă toate variantele caracteristicii sînt egale, ca de pildă :

$$x_1 = x_2 = x_3 = 5,$$

atunci,

$$\bar{x} = \frac{\sum_{i=1}^{3} x_i}{n} = \frac{15}{3} = 5,$$

adică $\bar{x} = x$.

2°. Suma abaterilor valorilor individuale de la media lor aritmetică, ponderat corespunzător, este egală cu zero :

$$\Sigma (x_i - \bar{x}) n_i = 0. \qquad (X,2.4)$$

Egalitatea rezultă imediat din formula mediei aritmetice. Referindu-ne tot la exemplul precedent avem distribuția abaterilor din tabelul 21.

Tabelul 21

x_i	$x_i - \bar{x}$	n_i	$(x_i - \bar{x}) n_i$	
8	−1,625	3	−4,875	−6,125
9	−0,625	2	−1,250	
11	+1,375	1	+1,375	+6,125
12	+2,375	2	+4,750	
Σ	—	8	0	0

Se constată că

$$\Sigma(x_i - \bar{x}) n_i = -6,125 + 6,125 = 0.$$

3°. Suma abaterilor de la media aritmetică ridicată la pătrat, este mai mică decît suma abaterilor de la oricare altă valoare A, ridicate la pătrat, adică

$$\Sigma (x_i - \bar{x})^2 n_i < \Sigma (x_i - A)^2 n_i. \qquad (X,2.5)$$

Dacă în distribuția de mai sus, vom considera $A = 8$ și $A' = 12$ și vom calcula abaterile ridicate la pătrat, comparativ cu abaterile de la media aritmetică ridicate la pătrat, vom obține tabelul 22.

Se vede din acest tabel că suma pătratelor abaterilor de la media aritmetică ponderate (coloana 4, ultimul rînd) este mai mică decît suma

pătratelor abaterilor de la $A < \bar{x}$ (coloana 7) și decît suma pătratelor abaterilor de la $A' > \bar{x}$ (coloana 10).

Tabelul 22

x	$x_i - \bar{x}$	$(x - \bar{x})^2$	$(x_i - \bar{x})^2 n_i$	$x - A$	$(x - A)^2$	$(x - A^2) n_i$	$x - A'$	$(x - A')^2$	$(x - A'^2) n_i$	n_i
1	2	3	4	5	6	7	8	9	10	11
8	−1,625	2,64	7,92	0	0	0	−4	16	48	3
9	−0,625	0,39	0,78	1	1	2	−3	9	18	2
11	+1,375	1,89	1,89	3	9	9	−1	1	1	1
12	+2,375	5,64	11,28	4	16	32	0	0	0	2
Σ	—	—	21,87	—	—	43	—	—	67	8

Expresia $\Sigma(x_i - A)^2 n_i$ din inegalitatea (X,2.5) poate fi considerată ca funcție de A, de pildă $\varphi(A)$, deoarece valoarea ei se schimbă în raport cu A. De aceea ne putem pune întrebarea: care trebuie să fie valoarea lui A pentru ca funcția $\varphi(A)$ să fie minimă? Condiția necesară și suficientă ca funcția $\varphi(A)$ să fie minimă la o valoare anumită a lui A este ca prima derivată a acestei funcții să fie nulă, iar derivata a doua — pentru aceeași valoare a lui A — să fie pozitivă.

Derivînd funcția $\varphi(A)$ avem:

$$\varphi'(A) = [\Sigma(x_i - A)^2 n_i]' = -2\Sigma(x_i - A)n_i . \qquad (X,2.6)$$

Pentru a afla valoarea lui A cînd funcția $\varphi(A)$ este minimă, anulăm derivata (X, 2,6):

$$-2\Sigma x_i n_i + 2\Sigma A n_i = 0, \qquad (X,2.7)$$

de unde deducem că

$$A\Sigma n_i = \Sigma x_i n_i$$

și deci

$$A = \frac{\Sigma x_i n_i}{n_i} = \bar{x} \qquad (X,2.8)$$

ceea ce înseamnă că numai în cazul cînd $A = \bar{x}$ vom avea într-adevăr suma abaterilor ridicate la pătrat minimă sau maximă. Să găsim acum derivata a doua a funcției $\varphi(A)$, adică

$$[-2\Sigma(x_i - A)\, n_i]' = 2\Sigma n_i . \qquad (X,2.9)$$

Se vede că $2\Sigma n_i$ — derivata a doua — este o valoare pozitivă.
Am demonstrat astfel, în general, proprietatea a treia arătată mai sus.

4°. Dacă toate variantele caracteristicii x sînt micșorate sau majorate de un număr de ori constant c, media se va micșora sau majora de același număr de ori. Astfel, dacă variantele seriei inițiale x_i le împărțim la numărul constant c, obținem o serie de noi variante

$$x_i' = \frac{x_i}{c} .$$

Prin urmare media $\bar{x}'$ va fi mai mică decît $\bar{x}$ de c ori.

Rezultă deci următoarea relație :

$$\bar{x} = \bar{x}' \cdot c$$

care poate fi folosită pentru calculul simplificat al mediei aritmetice.

Să considerăm exemplul din tabelul 23 în care variantele seriei inițiale le împărțim cu 100 :

Tabelul 23

x_i	n_i	$x : c = x'$	$x'_i n_i$	$\bar{x}'$	$\bar{x}$
1	2	3	4	5	6
500	6	5	30	$\bar{x}' = \frac{\Sigma x'_i n_i}{\Sigma n_i}$	$\bar{x} = \bar{x}' \cdot c = 8{,}45 \cdot 100$
700	2	7	14		
800	5	8	40		
1 000	4	10	40		
1 500	3	15	45		
Σ	20	—	169	8,45	845

Rezultatul obținut în coloana 6 se poate obține și direct cu ajutorul formulei

$$\bar{x} = \frac{\Sigma x_i n_i}{\Sigma n_i} = \frac{16\,900}{20} = 845.$$

5°. Dacă din variantele seriei inițiale scădem o valoare oarecare x_0, atunci media $\bar{x}'$ a valorilor astfel micșorate ($x_i - x_0 = x'_i$) va fi mai mică cu x_0 decît $\bar{x}$; cu alte cuvinte, vom avea relația

$$\bar{x} = \bar{x}' + x_0 . \qquad \text{(X,2.10)}$$

Într-adevăr, avem

$$\bar{x} = \frac{\Sigma n_i x_i}{\Sigma n_i}$$

$$\bar{x}' = \frac{\Sigma n_i x'_i}{\Sigma n_i} .$$

Însă, conform ipotezei,

$$x'_i = x - x_0 .$$

Prin urmare

$$\bar{x}' = \frac{\Sigma n_i (x_i - x_0)}{\Sigma n_i} = \frac{\Sigma n_i x_i}{\Sigma n_i} - x_0 = \bar{x} - x_0 .$$

Dacă în exemplul considerat la punctul 4°, vom lua $x_0 = 1\,000$, atunci calculul mediei va necesita operațiile din tabelul 24.

Tabelul 24

x_i	$x_i - x_0 = x'_i$	n_i	$x'_i \cdot n_i$	$\bar{x}' = \frac{\Sigma x'_i n_i}{\Sigma n_i} = \frac{-3\,100}{20} = -155$
1	2	3	4	
500	−500	6	−3 000	$\bar{x} = \bar{x}' + x_0 = -155 + 1\,000 = 845$ (rezultat care s-a obținut și prin metoda folosită în tabelul precedent).
700	−300	2	− 600	
800	−200	5	−1 000	
1 000 = x_0	0	4	0	
1 500	+500	3	+1 500	
Σ	—	20	−3 100	

6°. Dacă frecvențele n_i se împart la un număr constant c, media aritmetică a seriei nu se schimbă.

Dacă, de pildă, micșorăm de două ori frecvențele distribuției din tabelul 23 (vezi punctul 4° de mai sus), vom obține datele din tabelul 25.

Tabelul 25

x	$n' = \frac{n_i}{2}$	$x_i n'_i$	$\bar{x}' = \frac{\Sigma x_i n_i}{\Sigma n'_i} = \frac{8\,450}{10} = 845$
1	2	3	
500	3,0	1 500	
700	1,0	700	
800	2,5	2 000	Se vede că:
1 000	2,0	2 000	
1 500	1,5	2 250	$\bar{x}' = \bar{x}$
Σ	10,0	8 450	

7°. O consecință a proprietății de la punctul 6° este și proprietatea că media aritmetică poate fi calculată și pe baza frecvențelor relative care se obțin prin împărțirea frecvențelor absolute la numărul constant Σn_i (a se vedea formula (X,2.3)).

Dacă frecvența relativă o vom exprima în procente, atunci formula mediei aritmetice ponderate capătă forma:

$$\bar{w} = \sum x_i w_i \,, \qquad \text{(X,2.11)}$$

în care w_i este greutatea specifică, în procente, a fiecărei variante.

8°. Media aritmetică ponderată pentru întreaga colectivitate de variante este egală cu media mediilor de grupă ponderate cu frecvențele de grupă respective. Se spune, astfel, că media aritmetică admite o teoremă de adițiune.

Dacă avem mediile de grupă $\bar{x}_i$:

$$\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_n$$

și frecvențele de grupă respective

$$n_1, n_2, \ldots, n_n$$

media generală $\bar{x}$ este egală cu

$$\bar{x} = \frac{\Sigma \bar{x}_i n_i}{\Sigma n_i}.$$

Această proprietate se verifică ușor, pe baza datelor din tabelul 25. Vom împărți variantele caracteristice x în două grupe, prima cuprinde valorile 500, 700 și 800, cu frecvențele respective 3,1 și 2,5, iar a doua, valorile 1 000 și 1 500 cu frecvențele corespunzătoare 2 și 1,5.

Mediile de grupă $\bar{x}_i$ vor fi:

$$\bar{x}_1 = \frac{500 \cdot 3 + 700 \cdot 1 + 800 \cdot 2{,}5}{6{,}5} = \frac{4\,200}{6{,}5} = 646{,}5,$$

$$\bar{x}_2 = \frac{1000 \cdot 2 + 1500 \cdot 1{,}5}{3{,}5} = \frac{4\,250}{3{,}5} = 1\,214{,}3,$$

iar media generală $\bar{x}$:

$$\bar{x} = \frac{\Sigma \bar{x}_i n_i}{\Sigma n_i} = \frac{(646{,}5 \cdot 6{,}5) + (1\,214{,}3 \cdot 3{,}5)}{10} = \frac{8\,450}{10} = 845.$$

Combinînd diferite proprietăți ale mediei aritmetice putem ajunge la calculul ei simplificat, eliminînd o serie de operații greoaie ce se impun la calculul direct.

Vom da mai jos un exemplu de calcul, după ce vom expune și cîteva elemente ale teoriei momentelor.

În cazul unei distribuții continue, media aritmetică se calculează în mod analog

În intervalul (a,b) să considerăm funcția de frecvență $f(x)$. Notăm prin N volumul distribuției, adică:

$$\int_a^b f(x)\,dx = N.$$

Media aritmetică este definită prin relația

$$\xi = \frac{\int_a^b x f(x)\,dx}{N}. \qquad \text{(X, 2.12)}$$

În cazul unei distribuții discrete, am văzut că avem

$$\eta = \frac{\Sigma n_i x_i}{\Sigma n_i}. \qquad \text{(X,2.13)}$$

Prin urmare, media aritmetică în cazul unei distribuții discontinue se obține prin operația de însumare obișnuită (X,2.13) iar în cazul unei distribuții continue prin operația de integrare (X,2.12). Proprietățile 1°—8° pe care le-am arătat pentru media η se păstrează fără nici o modificare și pentru ξ. Acest lucru se explică prin faptul că formulele (X,2.12) și (X,2.13) pot fi înglobate într-una singură. Într-adevăr, atît ξ cît și η sînt cazuri particulare ale unei întregrale Stieltjes, care păstrează toate proprietățile 1°—8°, pe care am expus-o în capitolul VI.

În formula (X,2.12) presupunem că $f(x)$ reprezintă o funcție de frecvență relativă, ceea ce nu restrînge generalitatea.

Avem prin urmare:

$$\int_a^b f(x)\,dx = 1.$$

Notăm prin

$$F(x) = \int_a^x f(x)\,dx,$$

frecvența cumulată. Am văzut că pentru distribuțiile teoretice, $f(x)$ poartă numele de funcție de repartiție,

Frecvența $f(x)$ este definită în intervalul $[a,b]$. Luînd $f(x) = 0$, pentru $x < a$ și $x > b$, formulele de mai sus se pot scrie:

$$\int_{-\infty}^{+\infty} f(x)\,dx = 1, \qquad \text{(X,2.14)}$$

$$F(x) = \int_{-\infty}^{x} f(x)\,dx. \qquad \text{(X,2.15)}$$

În mod analog, introducem frecvența cumulată pentru o distribuție discretă prin formula

$$F(x_k) = \sum_{j=-\infty}^{k} f(x_j). \qquad \text{(X,2.16)}$$

În acest caz:

$$f(x) = f(x_j), \text{ pentru } x_j \leqslant x < x_{j+1}.$$

Reprezentată grafic, funcția $F(x)$ ne conduce la o curbă formată din trepte, cu salturile în punctele x_j.

Am definit astfel funcția de frecvență cumulată cu ajutorul densităților. Invers, să considerăm o funcție nedescrescătoare $F(x)$, continuă la dreapta, îndeplinind condițiile

$$0 \leqslant F(x) \leqslant 1,$$

$$F(-\infty) = 0, \quad F(+\infty) = 1.$$

Orice funcție $F(x)$, care îndeplinește condițiile de mai sus, poate fi considerată ca o funcție de frecvență cumulată, căreia îi corespunde o distribuție statistică mai generală, care conține, în anumite cazuri particulare, distribuția discretă și distribuția continuă. Gradul de generalitate al distribuției statistice definite cu ajutorul funcției $F(x)$ este ușor de determinat, ținînd seamă de modul cum am determinat gradul de generalitate al legilor de probabilitate definite prin funcția de repartiție (Cap. VI).

Scanned with OKEN Scanner

Să notăm prin $M(a,b)$ frecvența relativă pentru distribuția statistică definită de $F(x)$ și cuprinsă în intervalul $a < \xi \leqslant b$. Repetînd același raționament ca în cazul legilor de probabilitate, avem:

$$M(a < \xi \leqslant b) = F(b) - F(a),$$
$$M(a \leqslant \xi < b) = F(b - 0) - F(a - 0),$$
$$M(a < \xi < b) = F(b - 0) - F(a),$$
$$M(a \leqslant \xi \leqslant b) = F(b) - F(a - 0).$$

Frecvențele relative sînt astfel determinate în funcție de $F(x)$, pentru orice interval închis, deschis sau semideschis. Mai mult decît atît frecvențele relative sînt unic determinate, pentru $F(x)$ dat, pentru orice mulțime boreliană. Deci suportul pe care noi construim distribuțiile statistice sînt mulțimile boreliene.

Valoarea medie pentru o distribuție statistică se scrie:

$$\int_{-\infty}^{+\infty} x \mathrm{d}F(x).$$

Ea cuprinde drept cazuri particulare mediile date de formulele (X, 2.12) și (X, 2.13).

3. Media armonică

MEDIA ARMONICĂ SIMPLĂ. Dacă considerăm funcția determinantă de forma

$$\frac{1}{x_1} + \frac{1}{x_2} + \ldots + \frac{1}{x_n} = n \cdot \frac{1}{\overline{x}}, \qquad (X,3.1)$$

deducem imediat media armonică simplă

$$\overline{x}_n = \frac{n}{\sum \frac{1}{x_i}}. \qquad (X,3.2)$$

Vom nota media armonică cu $\overline{x}_n$ sau $\overline{x}_{(-1)}$; ea este un caz particular al mediei armonice ponderate.

MEDIA ARMONICĂ PONDERATĂ. Funcția determinantă pentru media armonică ponderată este de forma

$$\frac{1}{x_1} m_1 + \frac{1}{x_2} m_2 + \ldots + \frac{1}{x_n} m_n = \Sigma m_i \frac{1}{\overline{x}_h}, \qquad (X,3.3)$$

din care se deduce formula mediei armonice ponderate

$$\overline{x}_h = \frac{\sum_{i=1}^{n} m_i}{\sum_{i=1}^{n} \frac{1}{x} m_i}. \qquad (X,3.4)$$

Dacă considerăm din nou valorile:

$$8,\ 9,\ 11 \text{ şi } 12$$

cu ponderile respective

$$3,\ 2,\ 1 \text{ şi } 2$$

pentru care am arătat în paragraful precedent că valoarea mediei aritmetice este egală cu $\bar{x} = 9{,}625$, observăm că media armonică ponderată a acestora este mai mică decît media aritmetică:

$$\bar{x}_h = \frac{3+2+1+2}{\frac{1}{8}\cdot 3+\frac{1}{9}\cdot 2+\frac{1}{11}\cdot 1+\frac{1}{12}\cdot 2} = 9{,}345,$$

deci

$$\bar{x}_h < \bar{x}.$$

Din punct de vedere teoretic se demonstrează că media armonică este mai mică decît media aritmetică, dacă toate valorile x_i sînt pozitive *).

În studiul fenomenelor social-economice se pune deseori întrebarea cînd anume este necesar să se aplice media armonică. Sînt oare, între diferitele caracteristici, relaţii care pot lua forma funcţiei determinante de tipul (X,3.3) ? Răspunsul la această întrebare este desigur afirmativ.

Să considerăm exemplul din tabelul 26 în care avem datele privind productivitatea muncii a cinci muncitori, exprimată în unităţi naturale (w_i) şi sub forma consumului de timp de muncă pe unitate de produs (t_i) ca şi datele privind timpul total (T) lucrat în decursul unei zile de fiecare muncitor.

Tabelul 26

Muncitori	Numărul pieselor executate în timp de o oră (w_i)	Consumul specific de timp de muncă (t_i)	Numărul orelor lucrate (T_i)
Primul muncitor	5	0,20	7
al doilea ,,	4	0,25	6
al treilea ,,	4	0,25	8
al patrulea ,,	3	0,33	6
al cincilea ,,	6	0,16	8
	—	—	35

Pe baza datelor din tabelul 26 ne putem întreba, în primul rînd, care este productivitatea medie ($\bar{w}$) ce caracterizează acest colectiv de cinci muncitori.

Analiza economică a fenomenului complex al producţiei ne arată că volumul ei este determinat de doi factori: productivitatea muncii şi

*) Demonstraţia se găseşte la p. 67 şi următoarele.

timpul total lucrat; relația dintre volumul producției (Q), productivitatea muncii (w) și timpul total (T) se poate scrie sub forma

$$Q = \overline{w} \cdot T.$$

Pentru calculul mediei trebuie să avem funcția determinantă

$$\Sigma w_i T_i = \overline{w} \Sigma T_i. \qquad (X,3.5)$$

Folosind datele din tabelul 26, constatăm că media $\overline{w}$ este de forma mediei aritmetice și este egală cu :

$$\overline{w} = \frac{\Sigma w_i T_i}{T} = \frac{5 \cdot 7 + 4 \cdot 6 + 4 \cdot 8 + 3 \cdot 6 + 6 \cdot 8}{7 + 6 + 8 + 6 + 8} = \frac{157}{35} = 4,5.$$

În al doilea rînd, ne putem întreba care este consumul specific mediu de timp de muncă ($\overline{t}$) cu care lucrează colectivul nostru de muncitori. Este evident că lucrînd continuu cu acest consum specific mediu de timp de muncă, volumul producției trebuie să rămînă neschimbat.

În această condiție, vom avea următoarea relație determinantă :

$$Q = \Sigma \frac{T_i}{t_i} = \frac{\Sigma T_i}{\overline{t}} = \frac{1}{\overline{t}} \cdot \Sigma T_i. \qquad (X,3.6)$$

din care deducem că

$$\overline{t} = \frac{\Sigma T_i}{\Sigma \frac{1}{t_i} T_i}. \qquad (X,3.7)$$

Observăm că forma medie este cea a mediei armonice ponderate.

Folosind datele din exemplul nostru avem

$$\overline{t} = \frac{7 + 6 + 8 + 6 + 8}{\frac{1}{0,20} \cdot 7 + \frac{1}{0,25} \cdot 6 + \frac{1}{0,25} \cdot 8 + \frac{1}{0,33} \cdot 6 + \frac{1}{0,16} \cdot 8} = \frac{35}{157} = 0,22 \text{ ore.}$$

Rezultă că, în medie, un muncitor consumă 0,22 ore pentru a produce o piesă.

Trebuie să subliniem că între caracteristicile w și t există o relație inversă :

$$w = \frac{1}{t}; \qquad t = \frac{1}{w}$$

și că de obicei, cînd pentru una din ele funcția determinantă ne duce la forma mediei aritmetice, pentru cealaltă ne duce la forma mediei armonice.

Același lucru observăm și în exemplul nostru :

$$\overline{w} = \frac{157}{35}; \qquad \overline{t} = \frac{35}{157},$$

fiecare avînd sensul ei economic bine determinat.

Caracterul mediei armonice ca medie a mărimilor inverse ale caracteristicii considerate rezultă şi din aceea că formula mediei aritmetice a unei caracteristici oarecare x poate fi uşor transformată în formula mediei armonice şi invers :

$$\frac{\Sigma x}{n} = \frac{\Sigma x}{\Sigma x \cdot \frac{1}{x}} \quad \text{şi} \quad \frac{n}{\Sigma \frac{1}{x}} = \frac{\Sigma x \frac{1}{x}}{\frac{1}{x}}. \qquad \text{(X,3.8)}$$

Este de reţinut o concluzie de ordin metodologic de interes deosebit : problema formei matematice a mediei şi problema ponderilor folosite la construirea ei nu pot fi tratate izolat, rupte una de alta. Atît forma mediei cît şi sistemul anumit de ponderi trebuie să rezulte organic şi simultan din funcţia determinantă luată în considerare care, la rîndul ei, rezultă din analiza proprietăţilor esenţiale ale fenomenului economic luat în studiu.

În legătură cu această precizare, atragem atenţia, că dacă în exemplul din tabel ne-ar fi fost date : cantităţile (piese) produse de fiecare muncitor ($q_1 = 35$; $q_2 = 24$; $q_3 = 32$; $q_4 = 18$ şi $q_5 = 48$) în timpul său de lucru t_i şi T_i, atunci consumul specific mediu de timp de muncă ar trebui dedus pe baza unei alte funcţii determinante şi anume :

$$\Sigma T_i = \Sigma t_i q_i = \bar{t} \Sigma q_i \qquad \text{(X,3.9)}$$

din care se deduce uşor că

$$\bar{t} = \frac{\Sigma t_i q_i}{\Sigma q_i} = \frac{0{,}20 \cdot 35 + 0{,}25 \cdot 24 + 0{,}25 \cdot 32 + 0{,}33 \cdot 18 + 0{,}16 \cdot 48}{35 + 24 + 32 + 18 + 48} =$$

$$= \frac{35}{157} = 0{,}22 \text{ ore.}$$

După cum se vede, am obţinut acelaşi rezultat, însă, de data aceasta forma mediei este cea a mediei aritmetice ponderate, iar sistemul de ponderi este altul, ambele decurgînd în mod firesc dintr-o altă funcţie determinantă (timpul total lucrat şi nu volumul producţiei).

Calculul mediei armonice poate fi ilustrat şi grafic (vezi fig. 21).

Valorile date ale lui x se notează pe axa absciselor. Ridicînd linii perpendiculare în punctele respective de pe axa absciselor, pînă la intersecţia lor cu hiperbola caracteristicii inverse $y = \frac{1}{x}$ şi ducînd apoi orizontale (paralele cu axa absciselor) din aceste puncte de intersecţie pînă la axa ordonatelor obţinem pe aceasta din urmă axă, puncte corespunză-

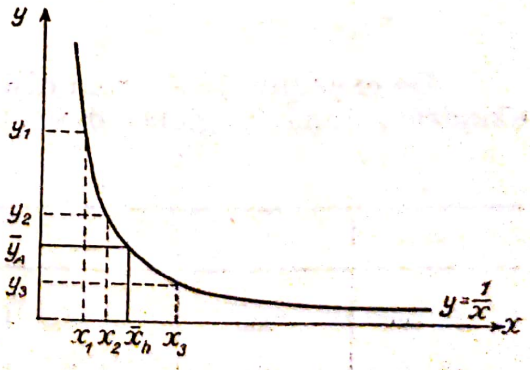


Fig. 21

toare pentru valorile inverse ale lui x. În continuare se calculează media aritmetică a valorilor lui y. Din punctul corespunzător mediei aritmetice $\bar{y}$ de pe axa coordonatelor ducem o linie orizontală pînă la intersecția ei cu hiperbola, iar de la acest punct de intersecție coborîm o linie perpendiculară pe axa absciselor. În punctul de intersecție a acestei perpendiculare cu axa absciselor se va afla valoarea medie armonică $\bar{x}_h$ a valorilor lui x. Punctul amintit de pe hiperbolă are coordonatele $\bar{x}_h$ și $\bar{y}_A$.

Prin urmare, valoarea găsită $\bar{x}_h$ este într-adevăr acea valoare a lui x care corespunde valorii mediei aritmetice a valorilor pe care le ia y, în concordanță cu relația invers proporțională care există între valorile celor două variabile.

4. Media pătratică

Pentru această medie sînt folosite notațiile M_2, $\bar{x}_2$, V_2, și altele, în funcție de caracteristica pentru care calculăm media pătratică.

MEDIA PĂTRATICĂ SIMPLĂ. Dacă avem funcția determinantă de forma

$$x_1^2 + x_2^2 + \ldots + x_n^2 = n\bar{x}_2^2 \qquad \text{(X,4.1)}$$

deducem media pătratică $\bar{x}_2$:

$$\bar{x}_2 = \sqrt{\frac{\sum_{i=1}^{n} x_i^2}{n}}. \qquad \text{(X,4.2)}$$

MEDIA PĂTRATICĂ PONDERATĂ. Funcția determinantă de forma

$$x_1^2 n_1 + x_2^2 n_2 + \ldots + x_n^2 n_n = \bar{x}^2 \Sigma n_i \qquad \text{(X,4.3)}$$

ne duce imediat la media pătratică ponderată:

$$\bar{x}_2 = \sqrt{\frac{\sum_{i=1}^{n} x_2^2 n_i}{\sum_{i=1}^{n} n_i}}. \qquad \text{(X,4.4)}$$

De exemplu dacă avem distribuția din tabelul 27 a stîlpilor de beton cilindrici, după lungimea diametrului mediu al bazei.

Tabelul 27

x_i	n_i	$\bar{x}^2$
15 cm	200	$\bar{x}_2 = \sqrt{\frac{\sum_{i=1}^{n} x_i^2 n_i}{\Sigma n_i}} = \sqrt{\frac{15^2 \cdot 200 + 18^2 \cdot 250 + 20^2 \cdot 300}{200 + 250 + 300}}$
18 cm	250	
20 cm	300	$= \sqrt{\frac{246\,000}{750}} = \sqrt{332{,}8} = 18{,}11$ cm

Valoarea medie a diametrului unui stîlp de beton din colectivitatea considerată este dată de media pătratică $\bar{x}_2$ care este egală cu 18,11 cm.

În studiul diferitelor elemente ale mijloacelor de producţie, se pun adeseori probleme la rezolvarea cărora trebuie să folosim media pătratică. Un exemplu clasic de aplicare a mediei pătratice este acela al calculului diametrului mediu al unor pietre de moară, ţinînd seama că funcţia determinantă (producţia totală a unei colectivităţi de mori cu diametre diferite ale pietrelor de moară) să rămînă neschimbată.

5. Media cubică

MEDIA CUBICĂ SIMPLĂ. Dacă avem funcţia determinantă de forma

$$x_1^3 + x_2^3 + \ldots + x_n^3 = \bar{x}_3^3 \cdot n, \qquad (X, 5.1)$$

atunci media cubică $\bar{x}_3$ este egală cu :

$$\bar{x}_3 = \sqrt[3]{\frac{\sum_{i=1}^{n} x_i^3}{n}}. \qquad (X,5.2)$$

Să presupunem că avem trei bile sferice cu razele de 1, 3 şi 4 cm, cu aceeaşi greutate specifică a masei egală cu 5. Ni se cere să aflăm raza medie, determinată astfel ca greutatea totală a celor trei bile să nu se schimbe (de pildă, să construim trei bile cu raze egale, avînd greutatea însumată a celor trei bile date). După cum se vede, greutatea totală va fi proprietatea determinantă care ne va ajuta să calculăm media cerută.

Se ştie că volumul sferei este egal cu $\frac{4\pi R^3}{3}$, iar greutatea ei este egală cu produsul dintre volum şi greutatea specifică a masei.

Prin urmare, greutatea totală a celor trei bile este egală cu

$$5 \cdot \frac{4}{3} \pi \cdot 1^3 + 5 \cdot \frac{4}{3} \pi \cdot 3^3 + 5 \cdot \frac{4}{3} \pi \cdot 4^3 = 5 \cdot \frac{4}{3} \pi (1^3 + 3^3 + 4^3).$$

Dacă cele trei raze ar fi egale între ele şi egale cu $\bar{x}$, atunci greutatea totală ar fi egală cu

$$5 \frac{4}{3} \pi (\bar{x}_1^3 + \bar{x}_2^3 + \bar{x}_3^3) = 5 \frac{4}{3} \pi \cdot 3 \bar{x}_3^3.$$

Funcţia determinantă (X, 5.1) ne arată că trebuie să avem egalitatea

$$5 \frac{4}{3} \pi (1^3 + 3^3 + 4^3) = 5 \cdot \frac{4}{3} \pi \cdot 3 \bar{x}_3^3$$

sau

$$1^3 + 3^3 + 4^3 = 3 \bar{x}_3^3$$

de unde deducem media cubică :

$$\bar{x}_3 = \sqrt[3]{\frac{1^3 + 3^3 + 4^3}{3}} = 3{,}13.$$

MEDIA CUBICĂ PONDERATĂ. Cînd valorilor x_i^3 li se asociază ponderile n_i, funcția determinantă are forma

$$\sum_{i=1}^{n} x_i^3 n_i = \bar{x}_3^3 \sum_{i=1}^{n} n_i \qquad (X,5.3)$$

din care deducem următoarea formulă a mediei cubice ponderate :

$$\bar{x}_3 = \sqrt[3]{\frac{\sum_{i=1}^{n} x_i^3 n_i}{\sum_{i=1}^{n} n_i}}. \qquad (X,5.4)$$

6. Media geometrică

MEDIA GEOMETRICĂ SIMPLĂ. Sînt uneori fenomene a căror caracteristică principală evoluează dinamic în ritm crescător, avînd funcția determinantă de forma produsului

$$x_1 \cdot x_2 \cdot x_3 \ldots x_n = \prod_{i=1}^{n} (x_i) = \bar{x} \cdot \bar{x} \ldots = \bar{x}^n \qquad (X,6.1)$$

din care deducem media geometrică simplă $\bar{x}_g$:

$$\bar{x}_g = \sqrt[n]{\prod_{i=1}^{n} (x_i)}. \qquad (X,6.2)$$

Dacă, de exemplu, creșterea productivității muncii pe trei perioade este caracterizată de următoarele date : 4%, 8% și 16% (cifrele reprezintă creșteri față de perioada precedentă); creșterea medie a productivității muncii se va calcula cu ajutorul formulei (X,6.2) :

$$\bar{x}_g = \sqrt[3]{1{,}04 \cdot 1{,}08 \cdot 1{,}16} = 1{,}092.$$

Pentru efectuarea rapidă a calculelor, expresia (X,6.2) se logaritmează, obținîndu-se astfel

$$\lg \bar{x}_g = \frac{\lg x_1 + \lg x_2 + \ldots + \lg x_n}{n} = \frac{\sum_{i=1}^{n} \lg x_i}{n}. \qquad (X,6.3)$$

Scanned with OKEN Scanner

Aplicînd această formulă la exemplul de mai sus, avem :

$$\lg \bar{x}_g = \frac{\lg 1{,}04 + \lg 1{,}08 + \lg 1{,}16}{3} = 0{,}038305,$$

din care deducem

$$\bar{x}_g = 1{,}092.$$

Atragem atenţia că, în general, media geometrică îşi găseşte aplicaţii în cazul indicilor dinamici. Deci, este vorba de serii în timp şi nu de distribuţii statice.

După cum vom vedea şi în capitolele consacrate analizei seriilor în timp, pentru calculul ritmului mediu de dezvoltare a diferitelor fenomene se foloseşte în practică şi media geometrică.

Dacă avem seria

$$y_0, y_1, y_2, \ldots, y_n,$$

formula mediei geometrice a indicilor în lanţ este

$$\bar{x}_g = \sqrt[n-1]{\frac{y_1}{y_0} \cdot \frac{y_2}{y_1} \ldots \frac{y_n}{y_{n-1}}} = \sqrt[n-1]{\frac{y_n}{y_0}} \qquad (X,6.4)$$

sau

$$\lg \bar{x}_g = \frac{\lg y_n - \lg y_0}{n-1}. \qquad (X,6.5)$$

MEDIA GEOMETRICĂ PONDERATĂ. Dacă notăm rapoartele din formula (X,6.4) cu i_j, fiecare repetîndu-se respectiv de $f_1, f_2, \ldots, f_n$ ori, atunci media indicilor individuali va fi dată de formula mediei geometrice ponderate :

$$\bar{x}_g = \sqrt[\Sigma f_j]{i_1^{f_1} \cdot i_2^{f_2} \cdot i_3^{f_3} \ldots i_n^{f_n}} = \sqrt[\Sigma f_j]{\prod_{j=1}^{n} (i_j^{f_j})}. \qquad (X,\ 6.6)$$

Pentru diferite aplicaţii ale mediei geometrice, a se vedea capitolul XIV. Tot acolo expunem şi metodologia de calcul a mediilor cronologice.

Trebuie să mai relevăm aici că media geometrică, ca şi media aritmetică, admite o teoremă de stabilitate. Media geometrică a unei serii formată din contopirea mai multor grupe ale căror medii geometrice sînt $\bar{x}_{g_1}, \bar{x}_{g_2}, \ldots, \bar{x}_{g_n}$, fiecare conţinînd $n_1, n_2, \ldots n_n$ elemente, este dată de formula

$$\lg \bar{x}_g = \frac{\sum_{i=1}^{n} n_i \lg \bar{x}_{g_i}}{\sum_{i=1}^{n} n_i} \qquad (X,6.7)$$

care devine

$$\lg \bar{x}_g = \frac{\sum_{i=1}^{n} \lg \bar{x}_{g_i}}{n} \qquad \text{(X,6.8)}$$

cînd grupele componente sînt formate din același număr de elemente.

7. Medii de ordin superior. Momente

Funcțiile determinante examinate pînă acum, care ne-au condus la formulele respective ale mediilor aritmetică, armonică, pătratică, cubică și geometrică, pot fi considerate ca fiind cazuri particulare ale unei formule mai generale, dedusă din funcția determinantă de forma următoare:

$$w = x^k; \qquad \text{(X,7.1)}$$

și întrucît avem

$$\bar{w} = \bar{x}_k \qquad \text{(X,7.2)}$$

și în același timp

$$\bar{w} = \frac{\sum_{i=1}^{n} w_i}{n} = \frac{\sum_{i=1}^{n} x_i^k}{n},$$

atunci

$$\bar{x}_k = \sqrt[k]{\bar{w}} = \sqrt[k]{\frac{\sum_{i=1}^{n} x_i^k}{n}} = \left[\frac{\sum_{i=1}^{n} x_i^k}{n}\right]^{\frac{1}{k}}. \qquad \text{(X,7.3)}$$

Dacă dăm lui k valori de la -1 la $+k^i$, satisfăcîndu-se inegalitățile $k < k' < k'' \ldots$, obținem șirul de valori medii simple de la media armonică $\bar{x}_h$ pentru care $k = -1$, apoi media geometrică pentru care $k = 0$, media aritmetică ($k = 1$) etc. Obținem, astfel, inegalitățile

$$\bar{x}_h < \bar{x}_g < \bar{x} < \bar{x}_2 < \bar{x}_3 < \ldots,$$

valabile cu condiția ca valorile $\bar{x}_i$ să fie pozitive.

Demonstrația teoretică a acestor inegalități dintre medii am dat-o la capitolul IV cînd ne-am ocupat în general de valorile medii ale unei variabile aleatoare.

Dacă luăm, de exemplu, valorile:

$$x_1 = 3,$$
$$x_2 = 5,$$
$$x_3 = 7,$$

calculul mediilor uzuale de care ne-am ocupat mai sus ne dă următoarele rezultate (folosind formulele mediilor simple):

$$\bar{x}_h = 4{,}44;\ \bar{x}_g = 4{,}72;\ \bar{x} = 5;\ \bar{x}_2 = 5{,}26 \text{ etc.}$$

Pentru media de ordin general, ponderată, funcția determinantă este

$$w = x^k n. \qquad (X,7.4)$$

Notînd cu

$$z = x^k,$$

avem

$$\overline{Z} = \frac{\sum_{i=1}^{n} z_i n_i}{\sum_{i=1}^{n} n_i} = \frac{\sum_{i=1}^{n} x_i^k n_i}{\sum_{i=1}^{n} n_i} \qquad (X,7.5)$$

și întrucît

$$\overline{Z} = \overline{x}^k \text{ și } \overline{x} = \sqrt[k]{\overline{Z}}$$

avem, definitiv,

$$\overline{x}_k = \sqrt[k]{\frac{\sum_{i=1}^{n} x_i^k n_i}{\sum_{i=1}^{n} n_i}} = \left[\frac{\sum_{i=1}^{n} x_i^k n_i}{\sum_{i=1}^{n} n_i}\right]^{\frac{1}{k}}. \qquad (X,7.6)$$

Dacă în exemplul de mai sus atribuim valorilor $x_1 = 3$; $x_2 = 5$ și $x_3 = 7$, ponderile $n_1 = 2$, $n_2 = 6$ și $n_3 = 4$, vom obține șirul mediilor ponderate crescătoare:

$$\overline{x}_h = \frac{2 + 6 + 4}{\frac{2}{3} + \frac{6}{5} + \frac{4}{7}} = 4{,}92,$$

$$\overline{x}_g = \sqrt[12]{3^2 \cdot 5^6 \cdot 7^4} = 5{,}14,$$

$$\overline{x} = \frac{3 \cdot 2 + 5 \cdot 6 + 7 \cdot 4}{2 + 6 + 4} = 5{,}33,$$

$$\overline{x}_2 = \frac{9 \cdot 2 + 25 \cdot 6 + 49 \cdot 4}{2 + 6 + 4} = 5{,}51$$

etc.

Expresiile (X,7.3) sau (X,7.6) poartă numele de *momente ale distribuției*. După cum s-a observat, momentul de ordinul întîi este chiar media aritmetică.

8. Momente inițiale. Momente obișnuite. Momente centrate

Momentele unei distribuții caracterizează distribuția variantelor unei caracteristici. Ele reprezintă medii aritmetice ale abaterilor — ridicate la diferite puteri de la un anumit punct ales ca origine.

În funcție de punctul de origine ales deosebim trei grupe de momente: momente inițiale, momente obișnuite și momente centrate. În cadrul fiecărei grupe există, evident, o infinitate de momente în funcție de puterea k la care este ridicată abaterea ($k \geqslant 0$).

MOMENTE INIȚIALE. Cînd originea arbitrară de calcul a abaterilor se ia în punctul $A = 0$, obținem momente inițiale de diferite ordine. Din punct de vedere practic interesează primele patru momente.

Astfel, notînd cu ν_i aceste momente, formulele lor generale capătă următoarele expresii:

Momentul de ordinul zero ($k = 0$)

$$\nu_0 = \frac{\Sigma x_i^0 n_i}{\Sigma n_i} = 1 \qquad \text{(X,8.1)}$$

Momentul de ordinul întîi ($k = 1$):

$$\nu_1 = \frac{\Sigma x_i n_i}{\Sigma n_i}. \qquad \text{(X,8.2)}$$

Mai departe, pentru $k = 2$, $k = 3$ și $k = 4$, avem:

$$\nu_2 = \frac{\Sigma x_i^2 n_i}{\Sigma n_i}; \qquad \text{(X,8.3)}$$

$$\nu_3 = \frac{\Sigma x_i^3 n_i}{\Sigma n_i}; \qquad \text{(X,8.4)}$$

$$\nu_4 = \frac{\Sigma x_i^4 n_i}{\Sigma n_i}. \qquad \text{(X,8.5)}$$

MOMENTE OBIȘNUITE. Dacă plasăm originea distribuției în oricare valoare luată arbitrar $A \neq 0 \neq \bar{x}$, atunci putem calcula momente obișnuite de diferite ordine.

Formula generală a unui moment obișnuit pe care-l notăm cu ν' are următoarea expresie:

$$\nu'_k = \frac{\Sigma (x_i - A)^k n_i}{\Sigma n_i}. \qquad \text{(X,8.6)}$$

În particular vom avea formulele primelor patru momente:

$$\nu'_0 = \frac{\Sigma (x_i - A)^0 n_i}{\Sigma n_i}, \qquad \text{(X,8.7)}$$

$$\nu'_1 = \frac{\Sigma (x_i - A) n_i}{\Sigma n_i}, \qquad \text{(X,8.8)}$$

$$\nu_2' = \frac{\Sigma(x_i - A)^2 n_i}{\Sigma n_i}, \qquad (X,8.9)$$

$$\nu_3' = \frac{\Sigma(x_i - A)^3 n_i}{\Sigma n_i}, \qquad (X,8.10)$$

$$\nu_4' = \frac{\Sigma(x_i - A)^4 n_i}{\Sigma n_i}. \qquad (X,8.11)$$

MOMENTE CENTRATE. Dacă abaterile sînt calculate față de media aritmetică, calculul momentelor centrate se face după următoarea formulă generală (notăm momentele centrate cu μ):

$$\mu_k = \frac{\Sigma(x_i - \bar{x})^k n_i}{\Sigma n_i}. \qquad (X,8.12)$$

În particular, vom avea formulele primelor trei momente:

$$\mu_0 = \frac{\Sigma(x_i - \bar{x})^0 n_i}{\Sigma n_i} = 1, \qquad (X,8.13)$$

$$\mu_1 = \frac{\Sigma(x_i - \bar{x}) n_i}{\Sigma n_i} = 0, \qquad (X,8.14)$$

$$\mu_2 = \frac{\Sigma(x_i - \bar{x})^2 n_i}{\Sigma n_i}. \qquad (X,8.15)$$

Momentul centrat de ordinul întîi ne aduce aminte de cunoscuta proprietate a mediei aritmetice pe care am formulat-o la punctul 1 al acestui capitol, iar momentul centrat de ordinul doi este cunoscutul indicator al variației, denumit dispersie pe care l-am mai notat și cu σ^2 sau $D^2(x)$ și asupra căruia vom mai reveni.

9. Formule de trecere de la momente obișnuite la momente centrate. Calculul mediei aritmetice prin metoda momentelor

Am observat că momentele obișnuite se pot calcula mai ușor decît momentele centrate, întrucît pentru simplificarea calculelor putem alege cea mai convenabilă origine arbitrară.

Deși din punctul de vedere al tehnicii de calcul este mai rațional să ne folosim de momentele obișnuite, după cum se știe însă, pentru analiza variației fenomenelor studiate este necesar calculul momentelor centrate. În acest scop s-au elaborat o serie de formule prin care momentele centrate se pot calcula cu ajutorul momentelor obișnuite.

Vom deduce formulele respective de calcul pentru primele momente centrate.

Observăm că formula generală a momentului obișnuit (X,8.6) poate fi exprimată și sub altă formă introducîndu-se media aritmetică $\bar{x}$ sub forma

$$A = \bar{x} + \alpha. \qquad (X,9.1)$$

Înlocuind în formula generală a momentului obișnuit pe A cu $\bar{x} + \alpha$, obținem expresia

$$\nu_k' = \frac{\Sigma[(x_i - \bar{x}) - \alpha]^k n_i}{\Sigma n_i}. \qquad (X,9.2)$$

Aplicînd formula binomului lui Newton, valoarea de la numărător se poate descompune în mai mulți termeni. Păstrînd alternanța semnelor și acordînd fiecărui termen același numitor comun Σn_i, putem scrie primii termeni sub forma:

$$\frac{\Sigma(x_i - \bar{x})^k n_i}{\Sigma n_i} - k\alpha \frac{\Sigma(x_i - \bar{x})^{k-1} n_i}{\Sigma n_i} + \frac{k(k-1)}{2!}\alpha^2 \frac{\Sigma(x_i - \bar{x})^{k-2} n_i}{\Sigma n_i} -$$

$$- \frac{k(k-1)(k-2)}{3!}\alpha^3 \frac{\Sigma(x_i - \bar{x})^{k-3} n_i}{\Sigma n_i} + \ldots$$

Observăm că în fiecare din termenii descompunerii de mai sus figurează media abaterilor de la media aritmetică de diferite ordine și de aceea o putem scrie sub o altă formă, introducînd notațiile μ_k pentru momente centrate.

Obținem astfel:

$$\nu_k' = \mu_k - k\alpha\mu_{k-1} + \frac{k(k-1)}{2!}\alpha^2\mu_{k-2} - \frac{k(k-1)(k-2)}{3!}\alpha^3\mu_{k-3} +$$

$$+ \frac{k(k-1)(k-2)(k-3)}{4!}\alpha^4\mu_{k-4}. + \ldots \qquad (X,9.3)$$

Această expresie constituie formula de legătură dintre momentele obișnuite și momentele centrate. Pe baza ei putem calcula orice moment centrat, exprimat prin formulele momentelor obișnuite care se calculează mai ușor.

De exemplu, să vedem care este valoarea momentului centrat de ordinul întîi.

Nu este greu de observat că $\alpha = -\nu_1'$. Într-adevăr, dacă dăm lui k valoarea 1, formula generală (X,9.3) capătă următoarea expresie:

$$\nu_1' = \mu_1 - \alpha\mu_0. \qquad (X,9.4)$$

Cum $\mu_1 = 0$ și $\mu_0 = 1$, expresia (X,9.4) se poate scrie:

$$\alpha = -\nu_1'. \qquad (X,9.5)$$

Amintindu-ne că am dedus formula generală (X,9.3) pornind de la substituirea (X,9.1) în care dacă înlocuim pe α cu valoarea lui $-\nu_1'$, obținem

$$A = \bar{x} - \nu_1'$$

de unde, în sfîrșit, deducem că

$$\bar{x} = A + \nu_1'. \qquad (X,9.6)$$

Această expresie ne ajută să calculăm media aritmetică folosind metoda simplificată a momentelor obișnuite.

Să deducem din expresia (X, 9.3) valoarea momentului centrat de ordinul doi:

Înlocuind pe $k = 2$, obținem

$$\nu_2' = \mu_2 - 2\alpha\mu_1 + \frac{2 \cdot 1}{1 \cdot 2}\alpha^2\mu_0. \qquad (X,9.7)$$

Cum $\mu_0 = 1$ și $\mu_1 = 0$, expresia (X,9.7) devine

$$\nu_2' = \mu_2 - \alpha^2$$

de unde

$$\mu_2 = \nu_2' - \alpha^2.$$

Înlocuind pe α cu egalul său $-\nu_1'$, obținem definitiv:

$$\mu_2 = \nu_2' - \nu_1'^2. \qquad (X,9.8)$$

Raționînd în mod analog, obținem și celelalte formule ale momentelor centrate, exprimate prin momente obișnuite.

Sintetizînd cele spuse pînă acum și făcînd calculele respective, ajungem la următoarele formule de legătură între momentele centrate și momentele obișnuite (primele șase momente):

$$\mu_0 = \nu_0' = 1, \qquad (X,9.9)$$

$$\mu_1 = \nu_1' - \nu_0' = 0, \qquad (X,9.10)$$

$$\mu_2 = \nu_2' - (\nu_1')^2, \ (\mu_2 = \sigma^2), \qquad (X,9.11)$$

$$\mu_3 = \nu_3' - 3\nu_2'\nu_1' + 2(\nu_1')^3, \qquad (X,9.12)$$

$$\mu_4 = \nu_4' - 4\nu_3'\nu_1' + 6\nu_2'(\nu_1')^2 - 3(\nu_1')^4, \qquad (X,9.13)$$

$$\mu_5 = \nu_5' - 5\nu_4'\nu_1' + 10\nu_3'(\nu_1')^2 - 10\nu_2'(\nu_1')^3 + 4(\nu_1')^5, \qquad (X,9.14)$$

$$\mu_6 = \nu_6' - 6\nu_5'\nu_1' + 15\nu_4'(\nu_1')^2 - 20\nu_3'(\nu_1')^3 + 15\nu_2'(\nu_1)^4 - 5(\nu_1')^6. \qquad (X,9.15)$$

Scanned with OKEN Scanner

Pentru verificarea momentelor centrate se folosesc următoarele formule:

$$\mu_3 = \nu_3' - 3\,\mu_2\,\nu_1' - (\nu_1')^3, \qquad (X,9.16)$$

$$\mu_4 = \nu_4' - 4\mu_3\,\nu_1' - 6\mu_2\,(\nu_1')^2 - (\nu_1')^4, \qquad (X,9.17)$$

$$\mu_5 = \nu_5' - 5\mu_4\,\nu_1' - 10\,\mu_3\,(\nu_1')^2 - 10\,\mu_2\,(\nu_1')^3 - (\nu_1')^5, \qquad (X,9.18)$$

$$\mu_6 = \nu_6' - 6\,\mu_5\,\nu_1' - 15\,\mu_4(\nu_1')^2 - 20\,\mu_3(\nu_1')^3 - 15\,\mu_2(\nu_1')^4 - (\nu_1')^6. \qquad (X,9.19)$$

Cînd momentele centrate se calculează pentru serii de variații grupate pe intervale, atunci se face ipoteza că elementele fiecărei grupe sînt concentrate la mijlocul intervalului. Este evident că o asemenea ipoteză duce în realitate la anumite erori care — după cum vom vedea — pot fi eliminate prin calculul corecțiilor lui Sheppard; în mod corespunzător se schimbă și formulele momentelor centrate de ordinul par.

Să considerăm următorul exemplu de calcul numeric al primelor patru momente obișnuite și centrate.

Datele convenționale de calcul privind distribuția unui număr de 100 muncitori după mărimea cîștigului orar în acord sînt cuprinse în tabelul 28.

Tabelul 28

Calculul indicatorilor necesari pentru determinarea momentelor obișnuite și centrate

Intervalele de grupare a caracteristicii x	frecvența n	Valorile centrale ale intervalelor	Abaterile de la A simplificate cu c ξ'	$\frac{x-A}{c}=\xi'$				Indicatorii de control pt. primele patru momente obișnuite	
				$\xi' n$	$(\xi')^2 n$	$(\xi')^3 n$	$(\xi')^4 n$	$\xi'+1=\xi''$	$(\xi'')^4 n$
1	2	3	4	5	6	7	8	9	10
7 — 9	8	8	−2	−16	32	−64	128	−1	8
9 — 11	16	10	−1	−16	16	−16	16	0	0
11 — 13	24	12	0	−32	—	−80	—	+1	24
13 — 15	50	14	+1	50	50	50	50	+2	800
15 — 17	2	16	+2	4	8	16	32	+3	162
				+54	—	+66	—	Σ	994
Σ	100		ξ'	+22	106	−14	226	ξ''	994

$A = 12$; $c = 2$; $\bar{x} = 12{,}44$; $\sigma^2 = \mu_2 = 1{,}0116$; $\sigma = \sqrt{1{,}0116} \simeq 1{,}005$

Datele din tabel fac posibil, în primul rînd, calculul momentelor obișnuite.

Folosind formulele respective, avem :

$$\nu_0' = \frac{\sum_{i=1}^{n} \xi_i'^0 n_i}{\sum_{i=1}^{n} n_i} = \frac{100}{100} = 1,$$

$$\nu_1' = \frac{\sum_{i=1}^{n} \xi_i' n_i}{\sum_{i=1}^{n} n_i} = \frac{22}{100} = 0{,}22,$$

$$\nu_2' = \frac{\sum_{i=1}^{n} \xi_i'^2 n_i}{\sum_{i=1}^{n} n_i} = \frac{106}{100} = 1{,}06,$$

$$\nu_3' = \frac{\sum_{i=1}^{n} \xi_i'^3 n_i}{\sum_{i=1}^{n} n_i} = \frac{-14}{100} = -0{,}14,$$

$$\nu_4' = \frac{\sum_{i=1}^{n} \xi_i'^4 n_i}{\sum_{i=1}^{n} n_i} = \frac{226}{100} = 2{,}26.$$

Pentru verificarea rezultatelor obținute vom aplica următoarea formulă *) :

$$\nu_4'' = \nu_0' + 4\,\nu_1' + 6\,\nu_2' + 4\,\nu_3' + \nu_4', \qquad (X, 9.20)$$

adică

$$\nu_4'' = 1 + 4 \cdot 0{,}22 + 6 \cdot 1{,}06 + 4 \cdot (-0{,}14) + 2{,}26 = 9{,}94,$$

în care ν_4'' este momentul obișnuit de ordinul patru al lui ξ'' (vezi coloanele 9 și 10 din tabelul 28).

Rezultatul obținut se verifică, calculînd pe ν_4'' direct ; avem

$$\nu_4'' = \frac{994}{100} = 9{,}94.$$

Prin urmare, calculul momentelor obișnuite este corect.

*) Această formulă de control este atribuită statisticianului suedez C.V. L. Charlier.

Folosind rezultatele obținute, vom calcula acum momentele centrate, pe baza formulelor cunoscute :

$$\mu_0 = \nu'_0 = 1,$$

$$\mu_1 = \nu'_1 - \nu'_0 = 1 - 1 = 0,$$

$$\mu_2 = \nu'_2 - (\nu'_1)^2 = 1{,}06 - (0{,}22)^2 = 1{,}0116,$$

$$\mu_3 = \nu'_3 - 3\nu'_2\nu'_1 + 2(\nu'_1)^3 = -0{,}14 - 0{,}6996 + 0{,}021296 =$$

$$= -0{,}8396 + 0{,}021296 = -0{,}818304,$$

$$\mu_4 = \nu'_4 - 4\nu'_3\nu'_1 + 6\nu'_2(\nu'_1)^2 - 3(\nu'_1)^4 = 2{,}26 + 0{,}1232 +$$

$$+ 0{,}307824 - 0{,}00702768 = 2{,}68399632 \simeq 2{,}68.$$

Folosind formulele de control al momentelor centrate, avem :

$$\mu_3 = \nu'_3 - 3\mu_2\nu'_1 - (\nu'_1)^3 = -0{,}14 - 0{,}667656 - 0{,}010648 = -0{,}818304.$$

$$\mu_4 = \nu'_4 - 4\mu_3\nu'_1 - 6\mu_2(\nu'_1)^2 - (\nu_1')^4 \simeq 2{,}68.$$

Rezultatele obținute fiind identice, se verifică astfel că și momentele centrate sînt calculate corect.

Pe baza datelor din tabelul 28 se poate calcula ușor și media aritmetică, prin metoda directă și prin metoda momentelor.

a) *Metoda directă.* Avem

$$\bar{x} = \frac{(8.8) + (10.16) + (12.24) + (14.50) + (16.2)}{100} = 12{,}44 .$$

b) *Metoda momentelor.* Avem

$$\bar{x} = A + \nu'_1 \cdot c = A + 0{,}22 \cdot 2 = 12 + 0{,}44 = 12{,}44 .$$

După cum se vede, rezultatele obținute sînt identice.

CORECȚIA SHEPPARD. În ceea ce privește calculul momentelor mai este de făcut o observație. În tabelul 28, distribuția de variație din coloana 1 am transformat-o, obținînd datele coloanei 3. Cu alte cuvinte, toate valorile dintr-un interval au fost înlocuite printr-un număr egal de elemente, egal — sub aspectul nivelului caracteristicii x — cu mijlocul intervalului. Or, prin această transformare se falsifică realitatea și este firesc ca valorile momentelor astfel calculate să fie afectate de anumite erori.

Corecțiile care trebuie aduse momentelor centrate brute pentru a obține momentele ajustate μ' sînt cunoscute sub numele de corecțiile lui Sheppard.

Dacă abaterile ξ_i au fost măsurate în intervalele distribuției, formulele de corecție pentru momente centrate sînt următoarele :

$$\left.\begin{aligned} \mu_1' &= 0, \\ \mu_2' &= \mu_2 - \frac{1}{12}, \\ \mu_3' &= \mu_3, \\ \mu_4' &= \mu_4 - \frac{\mu_2}{2} + \frac{7}{240}. \end{aligned}\right\} \qquad (X, 9{,}21)$$

Pentru momentele centrate exprimate în unități primare avem corecțiile :

$$\left.\begin{aligned} \mu_2' &= \mu_2 - \frac{1}{12}c^2, \\ \mu_4' &= \mu_4 - \frac{1}{2}\mu_2 c^2 + \frac{7}{240}c^4, \end{aligned}\right\} \qquad (X, 9{,}21a)$$

(în care c este mărimea intervalului).

Calculînd în exemplul nostru (tabelul 28) momentele centrate par corectate (după formula (X, 9, 21), obținem

$$\mu_2' = 1{,}0116 - 0{,}0833 = 0{,}9283$$

și

$$\mu_4' = 2{,}68 - 0{,}5058 + 0{,}02918 \simeq 2{,}20.$$

Corecțiile lui Sheppard, date în 1898, au fost formulate ulterior în mod mai precis, de statisticianul suedez H. Wold (1934).

Demonstrația formulelor Sheppard-Wold se face cu ajutorul cunoscutei formule de însumare Euler-Mac-Laurin din analiza matematică.

Să considerăm formula lui Euler-Mac-Laurin

$$\frac{1}{a}\int_{-\infty}^{+\infty} g(x)\,dx \int_{-\frac{a}{2}}^{\frac{a}{2}} f(x+u)\,du = \sum_{i=-\infty}^{+\infty} g(x_i) \int_{-\frac{a}{2}}^{+\frac{a}{2}} f(x_i+u)\,du + R. \qquad (X, 9.22)$$

Restul R din formula Euler-Mac-Laurin are o valoare bine cunoscută. Menționăm că, dacă în formula (X, 9.22) $g(x)$ este un polinom de gradul maxim k, atunci, în virtutea faptului că $x^k f^{(i)}(x)$, $(i = 1,2,\ldots,2s)$, este mărginită, rezultă după calcule simple

$$g(x)\int_{-\frac{a}{2}}^{\frac{a}{2}} f(x+u)\,du = 0\,(x^{-2}),$$

pentru $x \to \pm\infty$.

Cum restul R este de forma

$$R = (-1)^{s+1} a^{2s+1} \int_{-\infty}^{+\infty} G(x) \left[g(x_0 + ax) \int_{-\frac{a}{2}}^{\frac{a}{2}} f(x_0 + ax + u)\,du \right]^{2s+1} dx,$$

unde $G(x)$ este o funcție mărginită, oricare ar fi x, reiese, ținînd seama de valoarea asimptotică a lui x, că $|R| = O(a^{2s})$.

Dacă $g(x)$ este un polinom de gradul maxim k, caz care se va întîlni pentru rezolvarea problemei noastre, gradul de aproximație în formula lui Euler-Mac-Laurin este de ordinul a^{2s}. Vom reține acest lucru, neglijînd restul R în calculele care urmează:

Să luăm

$$\varphi(x) = \frac{1}{a}\int_{-\frac{a}{2}}^{\frac{a}{2}} g(x+u)\,du. \qquad \text{(X, 9.23)}$$

Derivînd, obținem

$$\varphi'(x) = \frac{1}{a}\int_{-\frac{a}{2}}^{\frac{a}{2}} g'(x+u)\,du = \frac{1}{a}\left[g\left(x+\frac{a}{2}\right) - g\left(x-\frac{a}{2}\right)\right].$$

În locul lui x, punem $x + \frac{a}{2}$ și egalitatea devine

$$\Delta_a g(x) = \varphi'\left(x+\frac{a}{2}\right), \qquad \text{(X, 9.24)}$$

unde

$$\Delta_a g(x) = \frac{g(x+a) - g(x)}{a}.$$

Integrînd prin părți în membrul întîi al formulei Euler-Mac-Laurin (X, 19.21) deducem

$$\frac{1}{a}\int_{-\infty}^{+\infty} g(x)\,dx\int_{-\frac{a}{2}}^{\frac{a}{2}} f(x+u)\,du = \left[\int_0^x g(t)\,dt\,\frac{1}{a}\int_{-\frac{a}{2}}^{\frac{a}{2}} f(x+u)\,du\right]_{-\infty}^{+\infty} -$$

$$-\frac{1}{a}\int_{-\infty}^{+\infty}\int_0^x g(t)\,dt\left[f\left(x+\frac{a}{2}\right) - f\left(x-\frac{a}{2}\right)\right]dx.$$

Dacă avem îndeplinită și condiția

$$\lim_{x\to\pm\infty}\left[\int_0^x g(t)\,dt\right]\left[\int_{-\frac{a}{2}}^{\frac{a}{2}} f(x+u)\,du\right] = 0, \qquad \text{(X, 9.5)}$$

atunci

$$\frac{1}{a}\int_{-\infty}^{+\infty} g(x)\,dx\int_{-\frac{a}{2}}^{\frac{a}{2}} f(x+u)\,du = -\frac{1}{a}\int_{-\infty}^{+\infty}\int_0^x g(t)\,dt\left[f\left(x+\frac{a}{2}\right) - f\left(x-\frac{a}{2}\right)\right]dx =$$

$$= -\frac{1}{a}\int_{-\infty}^{+\infty}\int_0^x g(t) f\left(x+\frac{a}{2}\right)dt\,dx + \frac{1}{a}\int_{-\infty}^{+\infty}\int_0^x g(t) f\left(x-\frac{a}{2}\right)dt\,dx$$

în prima parte a ultimei expresii facem schimbarea de variabilă $y = x + \frac{a}{2}$, iar în partea a doua, schimbarea de variabilă $y = x - \frac{a}{2}$.

Scanned with OKEN Scanner

Rezultă

$$\frac{1}{a}\int_{-\infty}^{+\infty} g(x)\,dx \int_{-\frac{a}{2}}^{\frac{a}{2}} f(x+u)\,du = -\frac{1}{a}\int_{-\infty}^{+\infty}\int_{0}^{y-\frac{a}{2}} g(t)\,f(y)\,dt\,dy +$$

$$+\frac{1}{a}\int_{-\infty}^{+\infty}\int_{0}^{y+\frac{a}{2}} g(t)\,f(y)\,dt\,dy = \frac{1}{a}\int_{-\infty}^{+\infty} f(y)\,dy \int_{y-\frac{a}{2}}^{y+\frac{a}{2}} g(t)\,dt.$$

Ținînd seama acum de relațiile (X, 9.22) și (X, 9.23) avem egalitatea

$$\int_{-\infty}^{+\infty} \varphi(x)\,f(x)\,dx = \sum g(x_i) \int_{-\frac{a}{2}}^{\frac{a}{2}} f(x_i+u)\,du. \qquad (X, 9.26)$$

Să alegem

$$\varphi(x) = x^m.$$

Pentru determinarea lui $g(x)$ trebuie să rezolvăm ecuația funcțională (X, 9.24)

$$\Delta_a\, g(x) = m\left(x+\frac{a}{2}\right)^{m-1}.$$

Soluția principală a acestei ecuații este

$$g(x) = B_m\left(\frac{x}{a}+\frac{1}{2}\right),$$

unde $B_m(x)$ este polinomul lui Bernoulli de gradul m.

În baza relației (X, 9.26) avem

$$\int_{-\infty}^{+\infty} x^m f(x)\,dx = \sum_{i=-\infty}^{+\infty} B_m\left(\frac{x_i}{a}+\frac{1}{2}\right)\int_{-\frac{a}{2}}^{\frac{a}{2}} f(x_i+u)\,du.$$

Știind că

$$B_m\left(\frac{x}{a}+\frac{1}{2}\right) = \sum_{i=0}^{m} C_m^i\, B_{m-i}\left(\frac{a}{2}\right)^{m-i} x^i.$$

B_{m-i} fiind numerele lui Bernoulli, rezultă

$$\mu_m = \sum_{i=0}^{m} C_m^i\, B_{m-i}\left(\frac{a}{2}\right)^{m-i} \mu'_{m'}, \qquad (X, 9.27)$$

deci formulele avute în vedere.

Dacă în formulele de mai sus facem $a = c$, găsim formulele de corecție ale lui Sheppard cu modificările aduse de H. Wold.

10. Medii de structură

Cele spuse pînă acum cu privire la valorile medii uzuale examinate nu ne dau însă nici o indicație în ceea ce privește posibilitățile de studiu — prin metode statistice — *a structurii* distribuțiilor statistice cu o singură caracteristică. Nici una din mediile examinate nu ne poate caracteriza *construcția* colectivității statistice luată în studiu. De aceea, din punct

de vedere teoretic și practic se pune problema de a fi determinate anumite valori ale caracteristicii care, prin poziția pe care o ocupă în volumul total al colectivității, fac posibilă caracterizarea structurii acesteia. Este evident că asemenea valori pot fi, de pildă, valoarea care se întîlnește cel mai des în cadrul colectivității studiate sau valoarea care ocupă locul central în distribuție, despărțind-o în două sau mai multe grupe egale ca mărime, ținîndu-se seama că toate variantele caracteristicii studiate sînt așezate într-o anumită ordine etc.

Tocmai aceste valori ale caracteristicii le vom considera ca fiind *valori medii de structură*, determinarea lor avînd un interes practic incontestabil.

Mediile de structură constituie un instrument prețios pentru a releva specificul diferitelor distribuții.

Le vom examina pe scurt.

1° MODUL (DOMINANTA). Prin definiție, *modul unei distribuții statistice este acea valoare a caracteristicii variabile căreia îi corespunde frecvența maximă.*

Dacă o serie discretă conține următoarele variante ale caracteristicii x_i cu frecvențele respective n_i :

x_i	1	4	7	10	7	4	1
n_i	5	10	15	20	15	10	5

atunci notînd cu M_0 valoarea modului, observăm că ea este egală cu 10, avînd frecvența maximă $n = 20$. Așadar, în distribuțiile discrete, valoarea modului se citește direct.

Pentru distribuții de variație cu intervale egale, valoarea-modul se determină cu ajutorul formulei :

$$M_0 = l_{M_0} + d \cdot \frac{\Delta_1}{\Delta_1 + \Delta_2}, \qquad (X, 10.1)$$

în care :

l_{M_0} este varianta de la limita inferioară a intervalului modal ;

d mărimea intervalului ;

Δ_1 diferența dintre frecvențele absolute sau relative, corespunzătoare intervalului modal și intervalului precedent ;

Δ_2 idem — corespunzătoare intervalului modal și intervalului următor.

Dacă ne referim la exemplul de la punctul precedent care ne-a servit pentru calculul diferitelor momente, valoarea-modul a cîștigului orar al colectivității de muncitori considerat se poate determina aplicînd formula (X, 10. 1).

Avem următorii indicatori necesari calculului :

$$l_{M_0} = 13 ; \; d = 2 ; \; n_{M_0} = 50 ; \; n_{M_0-1} = 24 ; \; n_{M_0+1} = 2,$$

$$\Delta_1 = 50 - 24 = 26 ; \; \Delta_2 = 50 - 2 = 48.$$

Prin urmare :

$$M_0 = 13 + 2 \cdot \frac{26}{26 + 48} = 13{,}702.$$

În cazul unei distribuţii continue, modul este abscisa corespunzătoare vîrfului curbei de distribuţie dacă această distribuţie are un singur vîrf determinant.

În aceste condiţii, valoarea modului reprezintă soluţia ecuaţiei

$$\frac{df(x)}{dx} = 0, \quad \left[\frac{d^2 f(x)}{dx^2} < 0\right], \qquad \text{(X, 10.2)}$$

unde presupunem că funcţia de frecvenţă $f(x)$ este continuă, împreună cu primele două derivate.

În practică se întrebuinţează procedee diferite pentru calculul cu aproximaţie al modului.

Ca o primă aproximaţie se consideră ca valoare a modului, valoarea variabilei din mijlocul intervalului corespunzător frecvenţei maxime (cu această aproximaţie s-a dedus şi formula de mai sus). Un alt procedeu care permite aflarea modului cu o aproximaţie satisfăcătoare constă în aplicarea formulei

$$M_0 = \frac{\Delta_1}{2n_{M_0} - n_{M_0+1} - n_{M_0-1}} \cdot \frac{d}{2}, \qquad \text{(X, 10.3)}$$

care este mai exactă decît formula (X, 10.1).

Arătăm pe scurt cum se determină formula aproximativă (X, 10.3).

Să considerăm intervalul de frecvenţă maximă a unei distribuţii continue, împreună cu intervalele de frecvenţă vecine lui :

$$\left(x_{M_0-1} - \frac{1}{2},\ x_{M_0-1} + \frac{1}{2}\right);\ \left(x_{M_0} - \frac{1}{2},\ x_{M_0} + \frac{1}{2}\right),\ \left(x_{M_0+1} - \frac{1}{2},\ x_{M_0+1} + \frac{1}{2}\right)$$

cărora le corespund frecvenţele :

$$n_{M_0-1} < n_{M_0} > n_{M_0+1}.$$

Ajustăm distribuţia în intervalul $\left(x_{M_0-1} - \frac{1}{2},\ x_{M_0+1} + \frac{1}{2}\right)$, cu ajutorul arcului de parabolă

$$y = a + bx + cx^2$$

ai cărui coeficienţi a, b şi c se pot determina din ecuaţiile

$$\int_{x_{M_0-1}-\frac{1}{2}}^{x_{M_0+1}+\frac{1}{2}} y\,dx = n_{M_0-1},\ \int_{x_{M_0}-\frac{1}{2}}^{x_{M_0}+\frac{1}{2}} y\,dx = n_{M_0},\ \int_{x_{M_0+1}-\frac{1}{2}}^{x_{M_0+1}+\frac{1}{2}} y\,dx = n_{M_0+1}. \qquad \text{(X, 10.4)}$$

Modul va fi dat de abscisa vîrfului parabolei a cărei valoare se obţine din ecuaţia

$$\frac{dy}{dx} = b + 2cx = 0. \qquad \text{(X, 10.5)}$$

Înlocuind acum în ecuaţia (X, 10.5) valorile coeficienţilor a, b şi c, date de ecuaţiile (X, 10.4) deducem formula (X,10.3).

Pentru distribuţiile uşor asimetrice, K. Pearson a stabilit următoarea formulă simplă pentru calculul modului

$$\bar{x} - 3(\bar{x} - M_e), \qquad \text{(X, 10.6)}$$

în care am notat prin $\bar{x}$ și M_e media aritmetică și respectiv mediana.

Trebuie să mai arătăm că există distribuții cu mai multe vîrfuri, numite *multimodale* (sau plurimodale) care au un *modul principal* și mai multe *module secundare*.

Metoda folosită pentru calculul modului principal este denumită *metoda grupărilor succesive*. Ea constă în regruparea distribuției multimodale, luîndu-se, în prima operație, intervale două cîte două, apoi trei cîte trei și așa mai departe. După cîteva regrupări succesive se ajunge la o distribuție *unimodală*, numărul modulurilor descrescînd pe măsura măririi intervalelor. Modulul principal al distribuției unimodale obținute astfel se calculează după una din formulele date mai sus, ținîndu-se seama de caracterul distribuției.

Modul ca valoare tipică de structură a distribuției se folosește atunci cînd ne interesează cea mai frecventă valoare, ca, de pildă, ora la care chemările telefonice sînt mai numeroase, ora de aglomerație a transportului în comun etc. cu scopul de a fi luate măsuri mai bune de ordin organizatoric.

În cazul unei distribuții simetrice, în mod teoretic media și mediana (vezi definiția medianei la punctul 2°) coincid cu modul. Coincidența acestor trei valori tipice ne dă un criteriu foarte bun pentru a preciza caracterul unei distribuții statistice. De foarte multe ori, în cazul distribuțiilor teoretice și modul coincide cu media aritmetică.

În legătură cu distribuțiile în formă de *U*, care am văzut că nu se întîlnesc în statistica social-economică, K. Pearson a introdus și noțiunea de *antimod* a cărui valoare corespunde frecvenței minime. Calculul antimodului se poate face după formula (X, 10.3).

2°. MEDIANA. Prin mediana unei distribuții se înțelege *acea valoare a caracteristicii care — toate unitățile colectivității fiind aranjate în ordinea de mărime a variantelor — împarte acest șir în două grupe egale ca număr.* Este vorba, deci, de o valoare într-adevăr mijlocie și o vom nota cu M_e.

Pentru determinarea practică a medianei sînt necesare cîteva precizări.

Dacă avem, de exemplu, o distribuție discretă și uniformă cu un volum de $2n + 1$ unități, mediana este reprezentată de elementul de rang $n + 1$. Dacă distribuția are un volum de $2n$ unități, pentru mediană se ia media aritmetică a elementelor de rangul n și $n + 1$.

De exemplu, dacă avem șiruri de valori descrescătoare :

$$65,\ 60,\ 55,\ 50,\ 45,\ 40,\ 35,$$

atunci valoarea mediană

$$M_e = 50,$$

dat fiind că 50% din numărul valorilor sînt de o parte a acestei valori centrale și 50% de altă parte.

Sau, dacă luăm același șir de valori în ordine inversă, adăugîndu-i încă una :

$$35,\ 40,\ 45,\ 50,\ 55,\ 60,\ 65,\ 70$$

atunci valoarea mediană va fi dată de media aritmetică a variantelor centrale

$$M_e = \frac{50 + 55}{2} = 52{,}5,$$

dat fiind că e vorba de un număr par de valori.

Formula generală care determină poziția medianei în cazul distribuțiilor discrete este deci următoarea :

$$\sum_{k \leqslant M_e} f_k = \sum_{k \geqslant M_e} f_k = \frac{1}{2}, \qquad \text{(X, 10.7)}$$

în care f_k reprezintă frecvențele relative. Este evident că în unele cazuri există un interval întreg în care putem alege în mod arbitrar mediana.

În cazul unei distribuții oarecare, *mediana este una dintre valorile ξ ale variabilei, pentru care avem*

$$\int_{-\infty}^{\xi} \mathrm{d}F(x) = \int_{\xi}^{+\infty} \mathrm{d}F(x) = \frac{1}{2}, \qquad \text{(X, 10.8)}$$

$F(x)$ fiind funcția de distribuție.

În cazul distribuției continue mediana este unic determinată de egalitatea

$$\int_{-\infty}^{\xi} f(x)\,\mathrm{d}x = \int_{\xi}^{+\infty} f(x)\,\mathrm{d}x = \frac{1}{2}, \quad [\mathrm{d}F(x) = f(x)\,\mathrm{d}x]. \qquad \text{(X, 10.9)}$$

Mediana se bucură de o proprietate deosebită care o face prețioasă în aplicații și anume :

Suma abaterilor în raport cu un punct de abscisă a este minimă, cînd această abscisă coincide cu valoarea medianei.

Luînd punctul de abscisă a și un alt punct arbitrar de abscisă b $(a < b)$, considerăm următoarele cazuri :

1) $x \leqslant a < b$. Atunci

$$\int_{-\infty}^{a} |x - a|\,\mathrm{d}F(x) = \int_{-\infty}^{a} (a - x)\,\mathrm{d}F(x) = \int_{-\infty}^{a} (b - x)\,\mathrm{d}F(x) - (b - a)\int_{-\infty}^{a} \mathrm{d}F(x). \qquad \text{(X, 10.10)}$$

2) $a \leqslant x \leqslant b$. Avem

$$\int_{a}^{b} |x - a|\,\mathrm{d}F(x) = \int_{a}^{b} (x - a)\,\mathrm{d}F(x) = -\int_{a}^{b} (b - x)\,\mathrm{d}F(x) + (b - a)\int_{a}^{b} \mathrm{d}F(x). \qquad \text{(X, 10.11)}$$

3) $a < b \leqslant x$. În acest caz,

$$\int_{b}^{+\infty} |x - a|\,\mathrm{d}F(x) = \int_{b}^{+\infty} (x - a)\,\mathrm{d}F(x) = \int_{b}^{+\infty} (x - b)\,\mathrm{d}F(x) + (b - a)\int_{b}^{+\infty} \mathrm{d}F(x). \qquad \text{(X, 10.12)}$$

Adunînd membru cu membru relațiile (X, 10.10), (X, 10.11) și (X, 10.12), rezultă

$$\int_{-\infty}^{+\infty} |x - a|\,\mathrm{d}F(x) = \int_{-\infty}^{+\infty} |x - b|\,\mathrm{d}F(x) - 2\int_{a}^{b} (b - x)\,\mathrm{d}F(x) - (b - a)\left[\int_{-\infty}^{a} \mathrm{d}F(x) - \int_{a}^{+\infty} \mathrm{d}F(x)\right]. \qquad \text{(X, 10.13)}$$

Scanned with OKEN Scanner

Dacă luăm acuma $a = M_e$ (valoarea medianei), egalitatea (X, 10.13) devine

$$\int_{-\infty}^{+\infty} |x - M_e| \, dF(x) = \int_{-\infty}^{+\infty} |x - b| \, dF(x) - 2\int_a^b (b - x) \, dF(x), \qquad (X, 10.14)$$

adică

$$\int_{-\infty}^{+\infty} |x - M_e| \, dF(x) < \int_{-\infty}^{+\infty} |x - b| \, dF(x). \qquad (X, 10.15)$$

Inegalitatea (X, 10. 15) ne arată tocmai proprietatea formulată mai sus, a medianei; se vede că suma abaterilor absolute ale variabilei în raport cu valoarea medianei este mai mică decît aceeaşi sumă a abaterilor, luate în raport cu abscisa oricărui alt punct.

Pentru determinarea cu aproximație a medianei se folosește următoarea formulă:

$$M_e = l_{M_e} + \frac{i\left(\frac{\Sigma n_i}{2} - S_n\right)}{n_{M_e}}, \qquad (X, 10. 16)$$

în care:

l_{M_e} este limita inferioară a intervalului median;

i mărimea intervalului median;

Σn_i suma frecvențelor (cumulată);

S_n suma frecvențelor tuturor intervalelor pînă la intervalul median; $S_n \leqslant \frac{\Sigma n_i}{2}$.

La baza acestei formule stă ipoteza că frecvențele se repartizează uniform în interiorul intervalului.

Să considerăm următorul exemplu de calcul al medianei:

Un colectiv de 40 muncitori se distribuie după procentul de depășire a nivelului mediu planificat al productivității muncii, după cum este arătat în tabelul 29.

Tabelul 29

Distribuirea unui colectiv de muncitori după procentul de depășire a nivelului mediu planificat al productivității muncii

Grupe de intervale ale procentului de depășire (x) %	Mijlocul intervalului (x_i)	Numărul muncitorilor n_i	Numărul cumulat al muncitorilor Σ_{n_i}
1	2	3	4
110—120	115	2	2
120—130	125	6	8
130—140	135	6	14
140—150	145	15	29
150—160	155	6	35
160—170	165	1	36
180—190	175	3	39
190—200	185	1	40

Jumătate din volumul acestei distribuții este 20, care corespunde unei valori din intervalul 135—145 de valori ale variabilei. În baza ipotezei repartiției uniforme a frecvențelor în fiecare interval, putem scrie pentru valoarea medianei :

$$M_e = 135 + \frac{10\left(\frac{40}{2} - 14\right)}{15} = 135 + \frac{60}{15} = 139.$$

Procentul de 139% față de nivelul mediu planificat al productivității muncii reprezintă valoarea mediană a distribuției considerate. Aceasta înseamnă că 50% din numărul muncitorilor depășesc, iar 50% nu depășesc acest nivel.

În cazul cînd valoarea medianei se poate lua cu o oarecare aproximație, putem folosi în acest scop reprezentarea grafică a funcției de distribuție $y = F(x)$. Obținem așa-numita *ogivă* a lui Galton, de care ne-am mai ocupat, luînd în sistemul coordonatelor rectangulare valorile caracteristicii ca abscise și ale frecvențelor cumulate ca ordonate.

Graficul unei ogive se prezintă ca o scară în trepte nedescrescătoare în cazul unei distribuții discrete (vezi fig. 22 a)) și ca o linie poligonală crescătoare, în cazul unei distribuții continue cu valorile variabliei grupate pe intervale egale (fig. 22 b)). Ordonata ultimei abscise va corespunde bineînțeles, frecvenței cumulate, egală cu volumul distribuției. Ducînd o paralelă la axa absciselor prin mijlocul acestei ordonate, tăiem ogiva într-un punct a cărui abscisă este mediana căutată.

Din studiul medianei se desprind unele observații interesante de care statisticianul trebuie să țină seama în rezolvarea problemelor practice. În primul rînd, mediana este o valoare tipică mult mai independentă

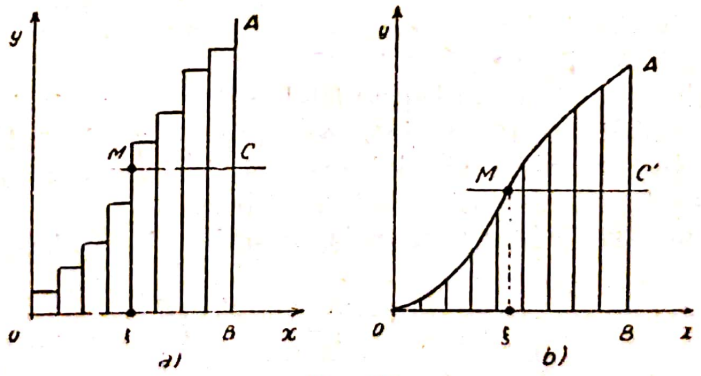


Fig. 22

decît media aritmetică, față de intervalele de grupare și forma distribuției. Din acest punct de vedere mediana este extrem de folositoare mai ales pentru distribuții deschise, unde limitele intervalului care conține mulțimea valorilor variabilei nu sînt fixate. În cazul unor asemenea distribuții media aritmetică conduce la valori mai puțin concludente.

În al doilea rînd, mediana, fiind mai puțin influențată de fiecare dintre elementele distribuției, poate fi utilizată cu mai multă încredere în cerce-

tările unde influența elementelor străine nu a putut fi înlăturată. Se spune astfel că mediana are o *stabilitate* mai mare decît media aritmetică.

Mediana prezintă însă și anumite inconveniente, ca de exemplu, lipsa teoremei de adițiune care după cum am văzut este valabilă pentru media aritmetică. Prin urmare, în cazul contopirii mai multor distribuții va trebui să calculăm din nou mediana distribuției obținute, medianele fiecărei distribuții în parte nedîndu-ne nici o indicație în această privință.

3°. QUANTILE. Din considerațiile făcute înainte am văzut că mediana, în unele împrejurări, ne poate da indicații mai bune despre o distribuție decît media. Utilizînd aceste calități ale ei, putem extinde noțiunea de mediană în vederea obținerii unor posibilități mai mari în studiul structurii distribuțiilor.

Astfel, ne putem propune de a găsi valorile ξ_i $(i = 1, 2, \ldots, n-1)$ ale variabilei pentru care avem

$$\int_{-\infty}^{\xi_1} \mathrm{d}F(x) = \int_{\xi_1}^{\xi_2} \mathrm{d}F(x) = \ldots = \int_{\xi_{n-1}}^{+\infty} \mathrm{d}F(x) = \frac{1}{n}, \qquad (X, 10.17)$$

$F(x)$ fiind funcția de distribuție.

Valorile ξ_i se numesc în general *quantile* și, după cum rezultă din egalitățile (X, 10.17), ele împart volumul N al colectivității în n părți egale. În aplicații, statistica întrebuințează unele valori particulare ale quantilelor.

Dacă $n = 4$, avem trei valori (Q_1, Q_2, Q_3) numite *quartile* care împart volumul distribuției în patru părți egale. Este evident că a doua quartilă, Q_2, coincide cu mediana M_e a distribuției.

Raportate la valoarea medianei, quartilele Q_2 și Q_3 se numesc respectiv *quartila mică* (inferioară) și *quartila mare* (superioară).

Un alt caz particular al quantilelor rezultă luînd $n = 10$. Cele nouă valori $(Q_1, Q_2, \ldots, Q_9)$, numite *decile*, împart volumul distribuției în zece părți egale. Decilele dau o măsură mai bună decît quartilele în ceea ce privește forma distribuției cercetate. Tot un caz particular al quantilelor sînt și *procentilele* $(n = 100)$.

Pentru calculul efectiv al quantilelor în general și al quartilelor, decilelor și procentilelor în special, se procedează la fel ca la calculul medianei, rămînînd valabile și aici precizările făcute acolo, atît pentru distribuțiile discrete cît și pentru cele continue.

De pildă, pentru prima și a treia quartilă, formulele de calcul sînt:

$$Q_1 = l_1 + \frac{\left(\Sigma n_i \frac{1}{4} - S_{Q_1}\right) d}{n_{Q_1}} \qquad (X, 10.18)$$

și

$$Q_3 = l_3 + \frac{\left(\Sigma n_i \frac{3}{4} - S_{Q_3}\right) d}{n_{Q_3}} \qquad (X, 10.19)$$

în care $S_{Q_1} \leqslant \Sigma n_i \frac{1}{4}$ și $S_{Q_3} \leqslant \Sigma n_i \frac{3}{4}$.

Exemple de calcul numeric al quartilelor sînt date în paragraful următor.

Trebuie să arătăm acum că valorile quartile şi decile sînt folosite şi pentru calculul unor indicatori ai variaţiei, pentru caracterizarea mai exactă a distribuţiilor studiate.

Pentru calculul quantilelor, în general, se poate recurge — la fel ca la valoarea mediană — şi la reprezentarea grafică, folosindu-se tot ogiva lui Galton despre care am vorbit mai sus. Desigur că se recurge la metoda grafică de determinare a quantilelor cînd nu este nevoie de o precizie prea mare. Astfel, quantila de rangul k va fi abscisa punctului de intersecţie a ogivei cu paralela la axa x-ilor, dusă prin diviziunea k a ordonatei corespunzătoare frecvenţei cumulate, egală cu volumul distribuţiei. Numărarea diviziunilor, egale între ele, se consideră pe ordonată de la axa x-ilor.

Dăm în figura 23 determinarea grafică a quartilelor şi decililor pentru exemplul din tabelul 29.

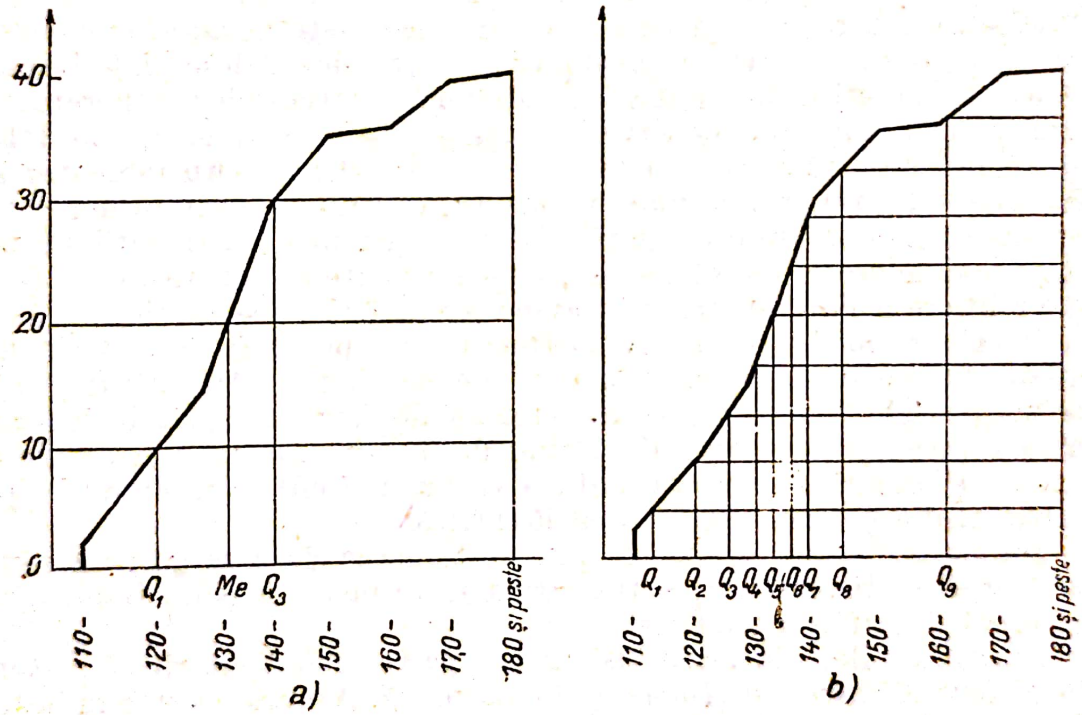


Fig. 23

Trebuie să precizăm acum că toate valorile medii prezentate se supun la două condiţii generale şi anume:

a) orice valoare medie trebuie să fie cuprinsă ca mărime între cea mai mică şi cea mai mare variantă a caracteristicii studiate;

b) valoarea mediei rămîne neschimbată dacă frecvenţele distribuţiei se împart sau se înmulţesc cu acelaşi număr.

Socotim util a mai desprinde o concluzie importantă ce decurge din condiţia *b*) de mai sus, pe care, de altfel, am formulat-o şi cu alt prilej

(vezi proprietățile mediei aritmetice): valorile tipice ale unei distribuții nu se schimbă, fiindcă sînt deduse fie pornind de la frecvențele absolute, fie de la frecvențele relative.

11. Indicatorii variației. Asimetrie

A. *Importanța statistică a noțiunii de variație sau împrăștiere.* Încă din capitolele consacrate studiului unei variabile aleatoare, pe plan teoretic, probabilistic, am atras atenția că valorile medii ale unei distribuții nu ne oferă decît o caracterizare numerică a unui singur aspect al acesteia. În general, o valoare medie nu ne oferă decît o imagine aproximativă, orientativă asupra caracterului distribuției studiate, rămînînd necunoscute trăsăturile ei esențiale. Pentru caracterizarea ei completă este necesară îndeosebi, cunoașterea organizării lăuntrice a distribuției, a modului cum valorile concrete — variantele caracteristicii — se repartizează în jurul valorii tipice și care poate diferi de la o distribuție la alta.

Practic, interesează în special cît de mare este *variația*, sau împrăștierea variantelor concrete în jurul mediei aritmetice. Adeseori, în lucrările teoretice și practice de statistică, abaterile variantelor concrete sînt denumite prin noțiunea generică de „*dispersie*". Noi, însă, am folosit și vom folosi această noțiune exclusiv pentru a denumi un indicator statistic, bine determinat și anume, momentul centrat de ordinul doi.

Importanța studiului științific al variației unei distribuții decurge din faptul că adeseori împrăștierea abaterilor poate să constituie un factor de diferențiere a două distribuții, avînd aceeași valoare medie.

O valoare medie, ca media aritmetică, nu poate fi un criteriu unic de diferențiere a două distribuții. Într-adevăr, două distribuții de același tip se pot deosebi fie numai prin valoarea medie, fie numai prin organizarea interioară (aceasta din urmă depinzînd de mărimea diferenței dintre cea mai mare și cea mai mică variantă a caracteristicii) sau, simultan, prin valoarea medie și prin organizarea lăuntrică.

De aceea, problema care se pune este aceea de a se găsi un sistem de indicatori statistici cu ajutorul căruia să se poată măsura *împrăștierea* caracteristicii statistic studiate.

Printre altele, din acest sistem fac parte și indicatorii ca: *abatere medie liniară, dispersie* și *abatere medie pătratică*. Asupra acestor indicatori nu vom mai insista însă aici, schemele lor teoretice de calcul ca și valoarea lor cognitivă, analitică, fiindu-ne cunoscute. Vom examina doar succint cîțiva alți indicatori ai variației ca și unele proprietăți ale dispersiei și ale abaterii medii pătratice (abaterii tip).

B. *Amploarea variației* (amplitudinea). Acest indicator ne dă o informație directă asupra gradului de ansamblu a variației, oglindind organizarea lăuntrică a distribuției. Îl vom nota cu R și-l vom calcula făcînd diferența, dintre cea mai mare (x_{max}) și cea mai mică (x_{min}) variantă a caracteristicii, exprimate în mărimi absolute:

$$R = x_{max} - x_{min}. \qquad (X, 11.1)$$

De pildă, pentru o serie de variabile cuprinzînd valorile 11, 8, 3, 2,6, 7, amplitudinea variației este egală cu 9, adică

$$R = 11 - 2 = 9.$$

Nu este greu de remarcat că acest indicator nu poate fi luat ca o măsură suficientă a împrăștierii, adică a modului de repartizare a variantelor caracteristicii în jurul uneia dintre valorile tipice, media aritmetică, mediana sau modul.

Se înțelege că în cazul distribuțiilor cu intervale de variație infinite, folosirea amplitudinii variației este lipsită de orice sens.

C. *Variația interquartilă.* După cum știm, 50% din unitățile unei serii sînt cuprinse între quartila întîi Q_1 și a treia Q_3 și anume 25% între Q_1 și M_e, iar 25% între M_e și Q_3. Este evident că cu cît valorile quartile se apropie de valoarea mediană, cu atît distribuția va prezenta o împrăștiere mai mică.

Rezultă, deci, că mărimea diferențelor dintre valorile quartile și mediana poate servi drept unul din indicatorii variației. Cea mai rațională construcție a acestui indicator — pe care-l vom nota cu Q_d — este dată de media aritmetică a intervalelor $Q_3 - M_e$ și $M_e - Q_1$, care în cazul distribuțiilor asimetrice sînt inegale. Prin urmare, *variația interquartilă* — — astfel vom denumi indicatorul în discuție — se poate determina cu ajutorul următoarei formule:

$$Q_d = \frac{(Q_3 - M_e) + (M_e - Q_1)}{2} = \frac{Q_3 - Q_1}{2}. \qquad (X, 11.2)$$

Avantajul acestui indicator al variației constă în faptul că se poate calcula chiar și pentru distribuțiile deschise.

Să considerăm următoarele exemple de distribuție:

x:	1	2	3	4	5	6,
n:	3	1	5	5	2	4.

Cele 20 de unități ale acestei distribuții se pot aranja sub forma unui șir pentru a fi scoase în evidență pozițiile valorilor quartile:

$$\underset{\uparrow\, Q_1}{1\ 1\ 1\ 2\ \overbrace{3\ \ 3}}\ \underset{\uparrow\, Q_2 = M_e}{3\ 3\ 3\ \overbrace{4\ \ 4}\ 4\ 4\ 4}\ \underset{\uparrow\, Q_3}{\overbrace{5\ \ 5}\ 6\ 6\ 6\ 6}$$

Avem

$$Q_1 = \frac{3+3}{2} = 3; \quad Q_2 = M_e = \frac{4+4}{2} = 4; \quad Q_3 = \frac{5+5}{2} = 5,$$

și atunci variația interquartilă Q_d este egală cu:

$$Q_d = \frac{Q_3 - Q_1}{2} = \frac{5-3}{2} = 1.$$

Acest indicator reprezintă de fapt variația celor 50% de variante din numărul total al variantelor — ce se grupează în jurul medianei distribuției.

Aceeași metodă de calcul se aplică și în cazul unei distribuții grupate pe intervale de variație.

Se determină mai întîi valorile quartilelor Q_1 și Q_3, conform formulelor (X, 10.18) și (X, 10.19) și apoi variația interquartilă se calculează, cu ajutorul formulei (X, 11.2).

Tabelul 30

Grupe de intervale (x)	Frecvența absolută (n)	Frecvența cumulată (Σn)
15,5	5	5
17,5	11	16
19,5	25	41
$l_1 \ldots \rightarrow$ 21,5	35 $\leftarrow n_{Q_1}$	76
23,5	40	116
$l_3 \ldots \rightarrow$ 25,5	36 $\leftarrow n_{Q_3}$	152
27,5	23	175
29,5	10	185
31,5	3	188

$\Sigma n = 188$

Dacă avem distribuția din tabelul 30, indicatorii necesari pentru calculul quartilelor Q_1 și Q_3 sînt — după cum știm — următorii:

$$l_1 = 21{,}5\,; \qquad \frac{\Sigma n}{4} = \frac{188}{4} = 47,$$

$$l_3 = 25{,}5\,; \qquad \frac{\Sigma n \cdot 3}{4} = (188.3) : 4 = 141,$$

$$S_{Q_1} = 41\,; \quad S_{Q_3} = 116\,; \quad n_{Q_1} = 35\,; \quad n_{Q_3} = 36,$$

$$d = 2.$$

Avem deci

$$Q_1 = 21{,}5 + \frac{(47 - 41)\cdot 2}{35} = 21{,}84,$$

$$Q_3 = 25{,}5 + \frac{(141 - 116)\cdot 2}{36} = 26{,}88.$$

Deducem imediat și indicatorul variației interquartile Q_d:

$$Q_d = \frac{26{,}88 - 21{,}84}{2} \simeq 2{,}5.$$

D. *Cîteva proprietăți ale dispersiei și ale abaterii medii pătratice.* După cum știm, dispersia, pe care am notat-o cu σ^2, μ_2 sau $D^2(x)$, repre-

zintă o măsură mai precisă a împrăştierii decît oricare alt indicator al variaţiei şi îndeosebi decît media abaterilor absolute. Trebuie să mai adăugăm acum că media abaterilor absolute nu poate fi folosită în problemele teoretice mai avansate, ca de exemplu, în problemele teoretice ale selecţiei datorită faptului că derivata ei este discontinuă în abscisele corespunzătoare frecvenţelor. De această defecţiune analitică nu suferă dispersia. Superioritatea dispersiei se vădeşte şi prin faptul că admite o teoremă de adiţie, ceea ce uşurează considerabil calculele cînd este vorba de contopirea mai multor distribuţii.

Pentru distribuţii continue, prin definiţie, dispersia este:

$$\sigma^2 = \int_{-\infty}^{+\infty} (x - \bar{x})^2 \, dF(x). \qquad (X, 11.3)$$

dacă membrul al doilea este finit.

Valoarea

$$\sigma = \sqrt{\mu_2} \qquad (X, 11.4)$$

ştim că este *abaterea medie pătratică*, care, după cum se vede, este o mărime de acelaşi ordin cu variabila x.

Aceşti indicatori sînt valori tipice pentru împrăştierea pe care o pun mai puternic în evidenţă decît media abaterilor absolute.

Dispersia se bucură de anumite proprietăţi interesante care sînt folosite la calculul ei:

a) *Dispersia este independentă faţă de schimbarea originii.* În capitolul precedent am văzut că momentele centrate de orice ordin sînt independente faţă de originea aleasă; dispersia fiind momentul centrat de ordinul al doilea proprietatea este demonstrată.

Avem

$$\sigma^2 = \frac{\Sigma[(x_i - A) - x_A]^2 n_i}{\Sigma n_i} = \frac{\Sigma(x_i - \bar{x})^2 n_i}{\Sigma n_i},$$

în care

$$\bar{x}_A = \frac{(x_1 - A) + (x_2 - A) + \ldots + (x_n - A)}{n} = \nu_1'.$$

b) *Dispersia este egală cu momentul obişnuit de ordinul doi, minus momentul obişnuit de ordinul întîi ridicat la pătrat:*

$$\sigma_x^2 = \nu_2' - (\nu_1')^2.$$

c) *Dacă toate variantele caracteristicii sînt înmulţite cu un număr constant k, dispersia se măreşte de k^2 ori, iar abaterea medie pătratică de k ori.*

De exemplu dacă toate variantele caracteristicii x_i le împărţim cu k, obţinem noi variante $\frac{x_i}{k} = z_i$ a căror dispersie σ_z^2 este

$$\sigma_z^2 = \frac{\Sigma(z_i - \bar{z})^2}{n} = \frac{\Sigma\left(\frac{x_i}{k} - \frac{\bar{x}}{k}\right)^2}{n} = \frac{1}{k^2} \cdot \frac{\Sigma(x_i - \bar{x})^2}{n} = \frac{\sigma_x^2}{k^2}. \qquad (X, 11.5)$$

Vom arăta, de asemenea, că la înmulțirea tuturor variantelor cu numărul constant k, abaterea medie pătratică se mărește cu acest număr în valoarea lui absolută. Astfel dacă avem noile variante $z = kx$, atunci abaterea medie pătratică a noilor variante este

$$\sigma_z = \sqrt{\frac{\Sigma(z - \bar{z})^2}{n}} = \sqrt{\frac{\Sigma(kx - k\bar{x})^2}{n}} = \sqrt{\frac{k^2 \Sigma(x - \bar{x})^2}{n}} =$$

$$= |k| \sqrt{\frac{\Sigma(x - \bar{x})^2}{n}} = |k| \sigma. \qquad (X, 11.6)$$

d) *Dispersia unei constante c este egală cu zero* :

$$\sigma_c^2 = 0,$$

întrucît toate variantele caracteristicii sînt egale între ele și egale cu media, iar toate abaterile sînt nule.

e) *Media aritmetică a abaterilor de la o constantă c, ridicată la pătrat, este mai mare decît dispersia caracteristicii x, diferența fiind egală cu pătratul diferenței dintre $\bar{x}$ și constanta c, adică*

$$\sigma_c^2 = \frac{\Sigma(x - c)^2}{n} = \frac{\Sigma(x - \bar{x})^2}{n} + (\bar{x} - c)^2. \qquad (X, 11.7)$$

f) *Pentru orice transformare liniară de forma $x = ay + b$, avem*

$$\sigma_x^2 = a^2 \sigma_y^2.$$

Plecăm de la egalitatea

$$\sigma_x^2 = \nu_2' - \nu_1^2.$$

Făcînd schimbarea de variabilă $x = ay + b$, deducem

$$\nu_2(ay + b) - [\nu_1(ay + b)]^2 = \nu_2(ay + b) - [a\nu_1(y) + b]^2 =$$
$$= a^2[\nu_2(y) - \nu_1^2(y)] = a^2\sigma^2(y)$$

și proprietatea este demonstrată.

g) *Dispersia unei distribuții care se poate descompune în mai multe distribuții parțiale este egală cu suma dispersiilor fiecărei distribuții în parte, adunată cu suma pătratelor abaterilor mediilor aritmetice ale distribuțiilor date, în raport cu media aritmetică a sumei distribuțiilor.*

Această proprietate este cunoscută sub numele de regulă de adunare a dispersiilor

Prin definiție putem scrie

$$\sigma^2 = \int_{-\infty}^{+\infty} (x - \bar{x})^2 dF(x) = \sum_{i=1}^{n} \int_{-\infty}^{+\infty} (x - \bar{x})^2 dF_i(x) = \sum_{i=1}^{n} \int_{-\infty}^{+\infty} (x - \bar{x}_i + \bar{x}_i - \bar{x})^2 dF_i(x) =$$

$$= \sum_{i=1}^{n} \int_{-\infty}^{+\infty} (x - \bar{x}_i)^2 dF_i(x) + \sum_{i=1}^{n} (\bar{x}_i - \bar{x})^2 = \sum_{i=1}^{n} \sigma_i^2 + \sum_{i=1}^{n} (\bar{x}_i - \bar{x})^2,$$

Scanned with OKEN Scanner

deoarece

$$\int_{-\infty}^{+\infty}(x-\bar{x}_{i_2})\mathrm{d}F_i(x)=0.$$

Teorema este prin urmare demonstrată.

Din proprietatea de mai sus rezultă următoarea consecință. Dispersia unei sume de distribuții este mai mare sau egală cu suma dispersiilor individuale. Egalitatea are loc dacă și numai dacă mediile distribuțiilor sînt egale între ele.

Prin urmare,

$$\sigma^2 > \sum_{i=1}^{n}\sigma_i^2, \text{ dacă } \bar{x}_{(i)} \neq \bar{x},$$

$$\sigma^2 = \sum_{i=1}^{n}\sigma_i^2, \text{ dacă } \bar{x}_{(i)} = \bar{x}.$$

h) *Media pătratelor tuturor diferențelor posibile ale perechilor de valori ale variabilei unei distribuții este egală cu dublul dispersiei.*

Această proprietate poate fi considerată ca o altă definiție a dispersiei.

Media diferențelor absolute ale valorilor variabilei unei distribuții este valoarea:

$$\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}|x-y|\,\mathrm{d}F(x)\mathrm{d}F(y). \qquad \text{(X, 11.8)}$$

Rezultă că media diferențelor absolute este media tuturor diferențelor absolute posibile ale perechilor de valori ale variabilei.

Expresia (X, 11.8) se scrie:

$$\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}|x-y|\,f(x)f(y)\mathrm{d}x\mathrm{d}y,$$

în cazul variabilei continue, și

$$\frac{1}{s^2}\sum_{i=-\infty}^{+\infty}\sum_{j=-\infty}^{+\infty}|x_i-x_j|\,s_is_j$$

sau

$$\frac{1}{s(s-1)}\sum_{i=-\infty}^{+\infty}\sum_{j=-\infty}^{+\infty}|x_i-x_j|\,s_is_j,$$

în cazul caracteristicii discontinue, respectiv pentru media diferențelor absolute cu repetiție sau fără repetiție a aceleiași valori a variabilei

$$(i \neq j,\ i=j \text{ sau } i \neq j).$$

Media diferențelor absolute a fost introdusă prima oară de C. Gini (1912) și are mai mult o însemnătate teoretică. Ea este independentă de orice valoare tipică și depinde numai de mulțimea valorilor variabilei. La fel ca și media abaterilor absolute, prezintă inconveniente de natură analitică, care o fac mai puțin proprie pentru probleme teoretice mai dificile.

Scanned with OKEN Scanner

E. *Coeficienții relativi ai variației. Asimetrie.* Pentru studiul comparativ a două distribuții de o singură caracteristică, un interes deosebit îl prezintă calculul coeficienților relativi ai variației.

În primul rînd, este util calculul *coeficientului de variație interquartil* raportîndu-se indicatorul variației interquartile (Q_d) la valoarea mediană a distribuției (M_e):

$$q = \frac{Q_d}{M_e} = \frac{\frac{Q_3 - Q_1}{2}}{M_e}$$

sau — ceea ce cu aproximație este același lucru

$$q = \frac{\frac{Q_3 - Q_1}{2}}{\frac{Q_3 + Q_1}{2}} = \frac{Q_3 - Q_1}{Q_3 + Q_1}. \qquad (X, 11.9)$$

Acest coeficient prezintă avantajul de a varia între două limite fixe -1 și $+1$. Trebuie arătat că acele cazuri în care aceste limite sînt atinse ($(Q_3 = 0)$ sau $(Q_1 = 0)$), nu prezintă nici o semnificație particulară, așa după cum se întîmplă, de exemplu, cu coeficientul de corelație (vezi capitolul XI).

În al doilea rînd, este deosebit de util calculul *coeficientului de variație* propus de K. Pearson.

Acest coeficient se poate calcula în două moduri:

$$V_\sigma = \frac{\sigma}{\bar{x}} \quad \text{sau} \quad V_\sigma = \frac{\sigma}{\bar{x}} \cdot 100 \qquad (X, 11.10)$$

în care se compară abaterea medie pătratică cu media aritmetică a distribuției sau

$$V_\xi = \frac{|\xi|}{\bar{x}} \quad \text{sau} \quad V_\xi = \frac{|\xi|}{\bar{x}} \cdot 100 \qquad (X, 11.11)$$

în care abaterea medie în valoare absolută (abaterea liniară) este aceea care se compară cu media aritmetică.

Forma unei distribuții poate fi apreciată și din punctul de vedere al *gradului de asimetrie* pe care-l prezintă. Gradul de asimetrie este un criteriu foarte bun pentru diferențierea distribuțiilor. Este evident că asimetria unei distribuții este cu atît mai pronunțată cu cît diferența dintre media aritmetică și modul este mai mare; cînd această diferență este nulă avem o indicație că sîntem în fața unei distribuții simetrice.

Pentru a avea o justă măsură a simetriei este util să raportăm diferența dintre media aritmetică și valoarea-modul, la abaterea medie pătratică:

$$a = \frac{\bar{x} - M_0}{\sigma}. \qquad (X, 11.12)$$

Este de notat că dacă $a > 0$, atunci valoarea-modul se află la stînga mediei ceea ce înseamnă că avem asimetrie de stînga; dacă $a < 0$, atunci valoarea-modul se află la dreapta mediei iar asimetria este de dreapta.

Acest coeficient a fost introdus tot de K. Pearson.

Un alt procedeu simplu de a măsura asimetria unei distribuții este acela de a se face raportul următor:

$$\alpha = \frac{Q_3 + Q_1 - 2M_e}{Q_d} \qquad \text{(X, 11.13)}$$

denumit coeficientul de asimetrie interquartil.

Se poate observa că cei doi coeficienți de asimetrie (X, 11.12) și (X, 11.13) sînt construiți cu ajutorul unor puncte de reper izolate (modul, mediana, media, quartilele), neexprimînd ansamblul abaterilor față de media aritmetică.

Or, numai un indicator care ar exprima tocmai acest raport ar fi cel mai potrivit pentru măsurarea asimetriei.

De aceea, pentru studiul asimetriei se folosesc și următorii indicatori ai lui K. Pearson:

$$\beta_1 = \frac{\mu_3^2}{\mu_2^3} \quad \text{și} \quad \beta_2 = \frac{\mu_4}{\mu_2^2} \qquad \text{(X, 11.14)}$$

în care μ_3 și μ_4 reprezintă momentele centrate de ordinul trei și patru.

Măsura relativă a asimetriei care trebuie socotită ca cea mai indicată pentru acest scop este

$$\sqrt{1} = \sqrt{\beta_1}.$$

Un ultim aspect intuitiv al formei unei distribuții este acela al *boltirii* pe care o prezintă.

Graficul din figura 24 ilustrează diferența — din punctul de vedere al boltirii — între două curbe de frecvențe simetrice, avînd aceeași dispersie.

Măsura boltirii este dată de coeficientul de ordinul doi, β_2, comparația făcîndu-se în raport cu curba normală, cînd $\beta_2 = 3$. Dacă avem inegalitatea

$$\beta_2 > 3,$$

curba este mai ascuțită spre vîrf decît curba normală și invers, dacă avem

$$\beta_2 < 3,$$

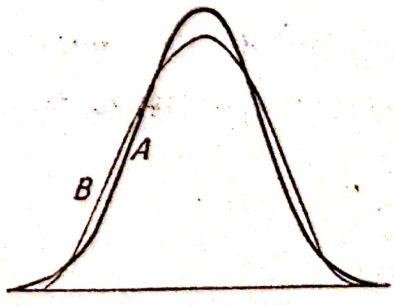


Fig. 24

curba este mai turtită spre vîrf decît o curbă normală.

Este util să mai reținem atenția asupra unui alt coeficient care indică gradul de îndepărtare a unei curbe de la curba normală, denumit *exces*. Acesta se calculează cu ajutorul formulei

$$E = \frac{\mu_4}{\sigma^4} - 3 \qquad \text{(X, 11.15)}$$

Dacă avem $\gamma_1 = 0$ și $E=0$, avem o indicație că curba se apropie de curba normală. Acești indicatori sînt, prin urmare, indicatori ai normalității distribuțiilor.

Aplicații ale tuturor coeficienților relativi de variație se pot face cu ușurință la exemplele de distribuții considerate pînă acum.

Am văzut multiplele aplicații ale momentelor la studiul unei distribuții statistice. De aceea se pune în mod legitim întrebarea dacă momentele caracterizează o distribuție statistică.

Mai precis, dacă ne dăm șirul infinit de numere M_i, $(i = 1, 2, \ldots \infty)$, definit prin relațiile :

$$M_i = \int_{-\infty}^{+\infty} x^i \mathrm{d}F(x)$$

putem determina funcția $F(x)$?

Această problemă este cunoscută în matematici sub numele de „problema momentelor". Primele contribuții importante pentru rezolvarea problemei au fost aduse de Cebîșev și Stieltjes.

Din punct de vedere matematic, problema momentelor preocupă și astăzi numeroși oameni de știință.

Din punct de vedere statistic, se poate afirma că dacă M_i reprezintă momentele unei distribuții statistice, adică există o funcție $F(x)$ îndeplinind condițiile unei funcții de frecvență cumulată, se pot găsi condițiile necesare și suficiente pentru ca această funcție să fie unică (Cebîșev).

De exemplu, dacă se dau relațiile

$$\frac{1}{\sqrt{\pi}} \int_{-\infty}^{+\infty} x^{2k} \mathrm{e}^{-x^2} \mathrm{d}x = \frac{1.3.5 \ldots (2k-1)}{2^k}$$

$$\frac{1}{\sqrt{\pi}} \int_{-\infty}^{+\infty} x^{2k-1} \mathrm{e}^{-x^2} \mathrm{d}x = 0,$$

în afară de distribuția a cărei densitate este $\frac{1}{\sqrt{\pi}} \mathrm{e}^{-x^2}$ nu mai există alta care să dea naștere la momentele de mai sus.

Scanned with OKEN Scanner

Capitolul XI

Distribuţii multidimensionale (cu mai multe caracteristici)

1. Consideraţii generale

Pînă acum am studiat distribuţii statistice referitoare la o singură caracteristică a unităţilor colectivităţii statistice, adică distribuţii unidimensionale. Există însă numeroase colectivităţi statistice ale căror unităţi prezintă două sau mai multe caracteristici. Pentru aceste colectivităţi frecvenţele se referă la mai multe caracteristici şi distribuţiile respective sînt multidimensionale.

Ca exemplu putem da o colectivitate formată dintr-un număr N de recruţi, pentru care s-a înregistrat înălţimea şi greutatea. Pentru fiecare unitate statistică avem două caracteristici : înălţimea şi greutatea. Statistica efectuată ne va da numărul recruţilor care au o înălţime dată şi o greutate dată sau a căror înălţime şi greutate variază în anumite intervale. Frecvenţa variază prin urmare în raport cu două caracteristici.

Este evident că studiul distribuţiilor multidimensionale înglobează ca un caz particular studiul distribuţiilor unidimensionale. În studiul distribuţiilor multidimensionale vom întîlni şi probleme noi, care nu au sens pentru distribuţiile unidimensionale. Ne referim astfel în primul rînd la problema corelaţiei. Fiind date două sau mai multe caracteristici se pune problema dacă aceste caracteristici sînt independente sau nu între ele.

Din exemplul pe care l-am dat s-ar putea trage bunăoară concluzia că sînt mai numeroşi tinerii înalţi care au greutatea de 80 kg, decît tinerii mărunţi care au aceeaşi greutate. Acesta ar fi un indiciu că între greutatea şi înălţimea unui om, făcînd parte dintr-o colectivitate, există o corelaţie adică o legătură de natură statistică. Se pune problema să se studieze în mod ştiinţific această corelaţie.

A doua problemă importantă este regresia unei caracteristici asupra celeilalte sau a mai multor caracteristici asupra altora. Dacă fixăm în exemplul nostru una din caracteristici, de exemplu greutatea, cealaltă caracteristică dă naștere la o distribuție unidimensională. Astfel toți tinerii cu greutatea de 80 kg, dau naștere la o distribuție unidimensională în raport cu înălțimea lor. Această distribuție are o medie aritmetică, mediană, dispersie etc. Se pune întrebarea cum sînt influențate aceste valori tipice, atunci cînd greutatea variază?

Ca și în cazul unei singure caracteristici, vom întîlni distribuții teoretice și distribuții empirice. Distribuțiile teoretice corespund unei legi de probabilitate date, cu două sau mai multe variabile. Masele corespunzătoare sînt probabilități.

După cum știm, distribuțiile empirice sînt rezultatul unei experiențe. Ele au un caracter statistic. De aceea le și numim distribuții statistice. Masele corespunzătoare sînt frecvențe.

Există un strîns paralelism între tratarea matematică a distribuțiilor teoretice și empirice. De foarte multe ori este suficient să se înlocuiască cuvîntul probabilitate prin frecvență pentru a trece de la o demonstrație privind o anumită proprietate a unei distribuții teoretice la demonstrația privind proprietatea analogă a unei distribuții empirice.

Legătura, dacă există, între o distribuție statistică și distribuția teoretică corespunzătoare este de aceeași natură ca în cazul unidimensional. Ea se bazează în mod esențial pe proprietățile de convergență pe care le-am examinat pe larg în capitolul VIII cînd am tratat legea numerelor mari.

În cazul unei singure dimensiuni, frecvența din punct de vedere matematic este o funcție de o singură variabilă. Din această cauză e ușor de reprezentat grafic, după cum am văzut, într-un spațiu cu două dimensiuni. Frecvențele unei distribuții statistice cu două dimensiuni sînt funcții de două variabile și deci reprezentarea lor grafică se poate face numai într-un spațiu cu trei dimensiuni. Reprezentarea devine mai dificilă și în același timp mai puțin utilă, fiind mult mai greu să se aprecieze pe baza unui desen caracteristicile unei suprafețe în spațiu decît ale unei curbe în plan. Dacă trecem, însă, la distribuțiile statistice cu mai mult de trei dimensiuni, ajungem la imposibilitatea realizării unei reprezentări grafice. O distribuție statistică cu trei caracteristici trebuie în mod normal să fie reprezentată ca o hipersuprafață într-un spațiu cu patru dimensiuni, ceea ce din punct de vedere grafic nu se poate realiza.

Asemenea dificultăți întîlnim nu numai la graficul distribuțiilor multidimensionale, ci și la studiul matematic al proprietăților lor. Pentru acest motiv, vom examina mai întîi distribuțiile bidimensionale, pentru care tratarea matematică este cu mult mai simplă. Pe de altă parte, este necesar să facem acest lucru și pentru faptul că în aplicații, cele mai frecvent întîlnite distribuții multidimensionale sînt cele cu două dimensiuni.

Proprietățile distribuțiilor cu mai mult de două dimensiuni vor fi arătate pentru cazul a trei dimensiuni, pentru a simplifica expunerea. Vom indica cazurile cînd pentru trecerea de la trei dimensiuni la un număr mai mare de dimensiuni se întîlnesc dificultăți speciale.

Scanned with OKEN Scanner

2. Elemente de analiză a distribuțiilor multidimensionale

În capitolul precedent am precizat deosebirile esențiale dintre funcțiile matematice care, uneori, reprezintă legi ale naturii — cum a fost legea căderii corpurilor în vid — pe de o parte și funcțiile experimentale, statistice, pe de altă parte. Am reținut atunci că corespondența $y = f(x)$ are caracterul unei funcții matematice riguros exacte, în sensul că y se consideră determinat odată cu x. Legătura dintre variabile este exprimată de obicei, într-o formulă analitică. În cazul distribuțiilor statistice în care își găsesc expresia legitățile statistice, legătura dintre variabila x și frecvențele respective nu are același caracter riguros, rigid și pentru aceeași valoare a primei variabile putem obține valori diferite pentru cea de-a doua, repetînd experiența de mai multe ori.

Această trăsătură distinctivă specifică a distribuțiilor statistice se vădește și mai mult în cazul distribuțiilor cu mai multe caracteristici. În acest caz, unei valori a primei din variabile îi corespunde o serie întreagă de valori ale celei de a doua, a treia ... variabile, distribuite într-un anumit mod. Se formează astfel o corespondență de un fel deosebit între valori și anume: nu de la valoare la valoare ci de la valoare la o întreagă distribuție sau mai multe distribuții. Prin urmare, ceea ce variază în raport cu una din variabile este structura corespunzătoare a altor variabile. Această variație este esențială pentru legitatea statistică sau, în limbajul statisticii matematice pentru o lege stohastică. În general, pentru distribuții cu mai multe caracteristici variabile, se calculează, ca și pentru distribuții cu o singură caracteristică, unele valori medii, momente de diferite ordine, cu ajutorul cărora le putem caracteriza. Se mai calculează însă și o serie de alți indicatori, specifici distribuțiilor cu mai multe caracteristici, cu ajutorul cărora se pot caracteriza multilateral legăturile reciproce dintre caracteristicele sistemului considerat.

Aspecte interesante prezintă și distribuțiile cu două caracteristici cu care începem studiul distribuțiilor multidimensionale.

* * *

O primă problemă care și-o pune un statistician ce studiază o distribuție cu două caracteristici este aceea de a o înscrie într-un tabel statistic și apoi de a o reprezenta grafic. Vom prezenta modul în care poate fi rezolvată această problemă, considerînd două caracteristici variabile x și y după care se distribuie cele 4 346 410 de familii din R.P.R.,*), caracteristica x reprezentînd numărul persoanelor care alcătuiesc o familie luînd valorile de la 2 la 9 și chiar peste 9, iar caracteristica y numărul persoanelor active din fiecare familie, luînd valorile de 0 la 8 și peste.

Distribuția este ilustrată în tabelul 31.

Datele din acest tabel se referă la întreaga țară, caracterizînd situația din ambele medii (urban și rural).

*) După datele recensămîntului din 21 februarie 1956.

Tabelul 31

Distribuţia familiilor din R.P.R. după numărul persoanelor care alcătuiesc familia şi după numărul persoanelor active din familie (Recensământul populaţiei din R.P.R. din 21 februarie 1956).

x \ y	Numărul persoanelor active în familie (y)									Total familii	%
	$0 = y_1$	$1 = y_2$	$2 = y_3$	$3 = y_4$	$4 = y_5$	$5 = y_6$	$6 = y_7$	$7 = y_8$	8 şi peste $= y_9$		
0	1	2	3	4	5	6	7	8	9	10	11
Numărul persoanelor care alcătuiesc o familie (x)											
$2 = x_1$	110 877	455 361	680 999	—	—	—	—	—	—	1 247 237	
$3 = x_2$	40 542	305 658	516 161	255 587	—	—	—	—	—	1 117 948	
$4 = x_3$	14 132	185 183	384 795	188 582	117 653	—	—	—	—	890 345	
$5 = x_4$	5 137	74 674	210 903	119 989	106 284	34 359	—	—	—	551 346	
$6 = x_5$	2 073	28 912	96 683	63 954	67 309	30 608	7 742	—	—	297 281	
$7 = x_6$	820	11 028	37 742	31 272	31 845	18 464	7 317	1 414	—	139 902	
$8 = x_7$	370	4 212	12 475	14 195	15 179	9 797	4 607	1 340	188	62 263	
9 şi peste $= x_8$	293	2 036	4 714	7 348	10 070	8 467	4 490	1 896	674	39 988	
Total	174 244	1 067 064	1 944 472	680 927	348 340	101 095	24 156	4 650	862	4 346 410	
%	4,0	24,6	44,7	15,7	8,0	2,3	0,6	0,1		100,0	

Scanned with OKEN Scanner

Citirea tabelului nu poate prezenta dificultăți. Astfel, dacă ne întrebăm cîte familii din R.P.R. sînt alcătuite din șase persoane din care patru sînt active? Răspunsul la această întrebare îl găsim imediat citind valoarea ce se află în rîndul x_5 la întretăiere cu coloana y_5. Se vede din tabel că numărul total al acestor familii în R.P.R. este de 67 309 *). Putem să ne punem și o altă întrebare: cîte familii alcătuite din cinci persoane sînt în R.P.R. indiferent cum se distribuie ele după numărul persoanelor active? Răspunsul îl găsim la întretăierea coloanei a 10-a cu rîndul x_4. Observăm că numărul acestor familii este de 551 346 ș.a.m.d.

Din examinarea tabelului 31 cititorul s-a putut convinge că este vorba de un tabel care ilustrează o distribuție discretă cu două caracteristici.

Putem considera și un alt tabel cu două caracteristici.

În tabelul 32 este dată distribuția unei grupe de 58 elevi în raport cu notele obținute la matematică și fizică.

Tabelul 32

Distribuția unui număr de elevi din două clase a VII-a în raport cu notele obținute la matematică și fizică.

x / y		Note obținute la matematică				
		10	9	8	7	Total
Note obținute la fizică	10	2	2	0	0	4
	9	1	6	1	0	8
	8	0	4	15	4	23
	7	0	0	10	13	23
	Σ	3	12	26	17	58

Pentru reprezentarea grafică a distribuțiilor de două caracteristici variabile în sistemul axelor rectangulare putem considera dreptunghiuri sau pătrate egale sau neegale între ele. În acest caz frecvențele corespunzătoare se referă la centrul acestor dreptunghiuri sau pătrate.

Construcții geometrice intuitive sau steriograme pentru astfel de distribuții, deși mai puțin folosite, se imaginează în spațiul cu trei dimensiuni.

Histograma distribuției este o steriogramă compusă din mai multe paralelipipede construite pe fiecare din dreptunghiuri sau pătrate și a căror înălțime este proporțională sau egală cu frecvențele corespunzătoare.

Reprezentarea geometrică cu ajutorul poliedrului frecvențelor și histogramei ne sugerează ideea, ca și la distribuțiile cu o singură caracteristică, de a considera distribuțiile cu două caracteristici ca prime aproximații ale unor distribuții teoretice care pot fi ele însile discontinue sau continue.

Distribuțiile cu două caracteristici pot fi reprezentate și prin cercuri de arie proporțională cu frecvențele distribuției.

*) Se înțelege că este vorba de situația din momentul recensămîntului din 21 februarie 1956.

Poziția acestor cercuri este fixată în raport cu două axe de coordonate corespunzînd valorilor respective ale celor două caracteristici. Dacă variația celor două caracteristici este dată de grupe de intervale, coordonatele cercului sînt determinate de mijlocul intervalelor corespunzătoare rubricii respective din tabelul de distribuții.

Tabelele statistice considerate pînă acum cuprind date statistice rezultate din observări directe și ordonate pe grupe de frecvențe.

După cum s-a mai arătat cu alt prilej, statistica nu studiază fenomenele de masă numai din punct de vedere descriptiv, ci caută să formuleze anumite legități generale, teoretice, în care aceste fenomene se încadrează.

Din acest punct de vedere fiecare rezultat al unei experiențe efective trebuie considerat ca una din valorile posibile cu *probabilitate dată*, ale unei caracteristici variabile.

Distribuția statistică este pusă astfel în legătură cu o variabilă aleatoare care este definită prin legea de probabilitate care stă la baza ei.

Rezultă, deci, că în tabelele sau sterogramele ilustrînd distribuții cu două caracteristici variabile, datele pot fi experimentale, sau teoretice, frecvențe sau probabilități propriu-zise.

A. *Legi de probabilitate cu două dimensiuni.* Să considerăm un vector $V(\xi, \eta)$, de componente ξ și η: $V(\xi, \eta)$ este o variabilă aleatoare vectorială, adică există o probabilitate corespunzătoare evenimentului

$$a_1 < \xi \leqslant a_2 \qquad b_1 < \eta \leqslant b_2.$$

pe care-l notăm sub formă vectorială

$$a < V \leqslant b,$$

unde a și b reprezintă vectorii $a(a_1\ a_2)$, $b(b_1,\ b_2)$.

Să notăm prin $P(a < V \leqslant b)$ această probabilitate și prin $F(x, y)$ funcția de două variabile reale x și y, definită prin egalitatea:

$$F(x, y) = P(\xi \leqslant x,\ \eta \leqslant y).$$

Funcția $F(x, y)$ poartă numele de funcție de repartiție. Ea se bucură de următoarele proprietăți:

$$0 \leqslant F(x, y) \leqslant 1,$$
$$F(-\infty, y) = 0, \quad F(x, -\infty) = 0,$$
$$F(+\infty, +\infty) = 1,$$
$$F(x + h, y + k) - F(x, y + k) - F(x + h, y) + F(x, y) \geqslant 0.$$

$F(x, y)$ este nedescrescătoare în raport cu fiecare variabilă x sau y luată în parte și continuă la dreapta.

Orice funcție $F(x, y)$ îndeplinind condițiile de mai sus este funcția de repartiție a unei unice legi de probabilitate bidimensionale.

$P(a < V \leqslant b)$ se exprimă cu ajutorul funcției de repartiție $F(x, y)$ în modul următor:

$$P(a < V \leqslant b) = F(b_1, b_2) - F(a_1, b_2) - F(b_1, a_2) + F(a_1, a_2).$$

În particular dacă avem:

$$\frac{\partial^2 F(x, y)}{\partial x\ \partial y} = f(x, y)$$

în orice punct (x, y), $f(x, y)$ fiind o funcție continuă de x și y, legea de probabilitate este continuă, iar $f(x, y)$ poartă numele de densitate de probabilitate.

B. *Distribuții statistice bidimensionale.* Distribuțiile statistice bidimensionale se definesc în același mod ca și legile de probabilitate bidimensionale, cu diferența că în loc de proba-

bilitatea $P(a < V \leqslant b)$, corespunzătoare evenimentului $a < V \leqslant b$, vorbim de $M(a < V \leqslant b)$, frecvența corespunzătoare intervalului cu două dimensiuni $a < V \leqslant b$ *). Se introduce funcția de frecvență cumulată sau ogiva $F(x, y)$ definită prin relația:

$$F(x, y) = M(\xi < x, \eta \leqslant y).$$

$F(x, y)$ se bucură de aceleași proprietăți ca și funcția de repartiție. Ea definește în mod unic o distribuție statistică bidimensională.

În cazul cînd

$$\frac{\partial^2 F(x, y)}{\partial x \partial y} = f(x, y)$$

în orice punct (x, y), $f(x, y)$ fiind continuă în raport cu x și y, distribuția statistică este continuă, iar $f(x, y)$ poartă numele de funcție de frecvență.

Momente ale distribuției de două caracteristici. Vom defini momentele unei distribuții de două caracteristici variabile luînd exemplul unei distribuții teoretice discrete, în care intervin deci probabilități. Considerațiile făcute aici sînt valabile și pentru distribuțiile statistice în care avem frecvențe.

Să considerăm vectorul aleator $V(x, y)$ care poate lua un număr finit sau o infinitate numărabilă de valori:

$$v_{kj} = v_{kj}(a_k, b_j), \qquad (k = 1, 2, \ldots; j = 1, 2, \ldots)$$

cu probabilități date $p_{kj} \geqslant 0$ astfel ca să avem

$$\sum_k \sum_j p_{kj} = 1. \qquad \text{(XI,2.1)}$$

Dacă $V(x, y)$ ia un număr finit de valori se poate folosi tabelul 33, care scoate în evidență valorile v_{kj} și probabilitățile respective **).

Tabelul 33

x \ y	b_1	b_2	...	b_j	...	b_n	
a_1	p_{11}	p_{12}	...	p_{1j}	...	p_{1n}	$p_{1\cdot}$
a_2	p_{21}	p_{22}	...	p_{2j}	...	p_{2n}	$p_{2\cdot}$
⋮							
a_k	p_{k1}	p_{k2}	...	p_{kj}	...	p_{kn}	$p_{k\cdot}$
⋮							
a_m	p_{m1}	p_{m2}	...	p_{mj}	...	p_{mn}	$p_{m\cdot}$
Σ	$p_{\cdot 1}$	$p_{\cdot 2}$		$p_{\cdot j}$		$p_{\cdot n}$	1

*) Mulțimea punctelor din plan, îndeplinind condițiile $a_1 < \xi \leqslant a_2$, $b_1 < \eta \leqslant b_2$, poartă numele de interval cu două dimensiuni.

**) A se vedea capitolul IV în care am examinat, printre altele, și determinarea valorilor medii pentru *sisteme* de un număr finit de variabile aleatoare.

Scanned with OKEN Scanner

După cum se vede, în tabel sînt m linii și n coloane. În total sînt $m + n$ alinieri, înțelegînd prin cuvîntul aliniere, o linie sau o coloană.

Am notat:

$$\left.\begin{aligned} \sum_{j=1}^{n} p_{kj} &= p_{k\cdot} \qquad (k = 1, 2, \ldots, m), \\ \sum_{k=1}^{m} p_{kj} &= p_{\cdot j} \qquad (j = 1, 2, \ldots, n). \end{aligned}\right\} \qquad \text{(XI, 2.2)}$$

Rezultă

$$\sum_{k=1}^{n} p_{k\cdot} = 1, \qquad \sum_{j=1}^{m} p_{\cdot j} = 1$$

$p_{k\cdot}$ și $p_{\cdot j}$ se numesc probabilitățile marginale și anume: $p_{k\cdot}$ reprezintă probabilitatea ca $x = a_k$, oricare ar fi valoarea pe care o va lua y, $p_{\cdot j}$ reprezintă probabilitatea egalității $y = b_j$, oricare ar fi x, prin x și y notînd componentele vectorului aleator $V(x, y)$.

Să notăm prin $p(x = a_k / y = b_j)$ probabilitatea condiționată a egalității $x = a_k$ știind că $y = b_j$.

Avem

$$p(x = a_k \mid y = b_j) = \frac{p(x = a_k, y = b_j)}{p(y = b_j)} = \frac{p_{kj}}{p_{\cdot j}} \qquad \text{(XI, 2.3)}$$

și în mod analog,

$$p(y = b_j \mid x = a_k) = \frac{p(x = a_k, y = b_j)}{p(x = a_k)} = \frac{p_{kj}}{p_{k\cdot}}. \qquad \text{(XI, 2.4)}$$

Și acum vom sublinia din nou următoarele:

Dacă evenimentele $x = a_k$ și $y = b_j$ sînt independente, avem pentru orice a_k și b_j egalitatea:

$$p(x = a_k, y = b_j) = p(x = a_k) \cdot p(y = b_j) = p_{k\cdot} \cdot p_{\cdot j}$$

sau

$$p_{kj} = p_{k\cdot} \cdot p_{\cdot j}. \qquad \text{(XI, 2.5)}$$

Reciproc, dacă această relație are loc pentru oricare $k = 1, 2, \ldots, m$, și $j = 1, 2, \ldots, n$ rezultă că cele două componente x și y ale vectorului V sînt independente între ele.

Vom aminti acum că dacă este îndeplinită egalitatea (XI, 2.5), atunci valoarea medie a produsului a două variabile independente dintr-un sistem este egală cu produsul valorilor medii ale variabilelor independente:

$$\overline{x \cdot y} = \bar{x} \cdot \bar{y}. \qquad \text{(XI, 2.6)}$$

Aceeași proprietate se păstrează și în general. Dăm în continuare două exemple de distribuții cu două caracteristici variabile.

EXEMPLUL 33. Într-o urnă sînt 50 de bile albe, 25 de bile roșii și 25 de bile negre, identice în afară de culoare. Se fac din urnă patru

extracții succesive, înapoindu-se de fiecare dată bila extrasă în urnă. Notăm cu x numărul de bile albe și cu y numărul de bile roșii ieșite din același număr de extracții. Este evident că x și y nu pot lua decît valorile 0, 1, 2, 3, 4.

Probabilitatea de a obține $x = a$, $y = b$, este egală cu

$$\frac{4!}{a!\,b!\,(4-a-b)!} \cdot \frac{1}{2^{8-a}}$$

în conformitate cu formula arătată la distribuția binomială.

În baza acestor date am întocmit tabelul 34.

Tabelul 34

	Numărul de bile roșii ieșite în patru extracții succesive						
	x \ y	0	1	2	3	4	
Numărul de bile albe ieșite în patru extracții succesive	0	c*)	4c	6c	4c		15c
	1	8c	24c	24c	8c	0	64c
	2	24c	48c	24c	0	0	96c
	3	32c	32c	0	0	0	64c
	4	16c	0	0	0	0	16c
	Total	81c	108c	54c	12c		1

*) În acest tabel s-a notat $c = \frac{1}{256}$ — pentru scriere mai simplă.

EXEMPLUL 34. Într-o urnă sînt șase bile albe și nouă bile negre. Se fac din urnă șase extracții succesive, fără a pune înapoi bila extrasă. Notăm cu x numărul de bile albe obținut în primele trei extracții și cu y numărul de bile albe obținut din ultimele trei extracții.

Avem tabelul 35 al distribuției.

Probabilitățile p_{ij} sînt calculate cu ajutorul formulelor stabilite la ribuția hipergeometrică.

Tabelul 35

x \ y	0	1	2	3	
0	0,016 79	0,075 52	0,075 52	0,016 79	0,184 62
1	0,075 52	0,226 58	0,151 04	0,021 57	0,474 71
2	0,075 52	0,151 04	0,064 73	0,005 39	0,296 68
3	0,016 79	0,021 57	0,005 39	0,000 20	0,043 95
	0,184 62	0,474 71	0,296 68	0,043 95	1

Este de observat simetria datelor din acest tabel.

Scanned with OKEN Scanner

În cazul general, să admitem că în urnă sînt A bile albe și N bile negre că în primele $\alpha + \beta$ extracții am obținut $x = \alpha$ bile albe și în următoarele $\alpha' + \beta'$ extracții $y = \alpha'$ bile albe. Probabilitatea ca să avem $x = \alpha$, $y = \alpha'$ este

$$\frac{C_A^{\alpha} \cdot C_N^{\beta}}{C_{A+N}^{\alpha+\beta}} \cdot \frac{C_{A-\alpha}^{\alpha'} \cdot C_{N-\beta}^{\beta'}}{C_{A+N-\alpha-\beta}^{\alpha'+\beta'}} = \frac{A!\,N!\,(\alpha+\beta)!\,(A+N-\alpha-\beta-\alpha'-\beta')!\,(\alpha'+\beta')!}{\alpha!\,\beta!\,\alpha'!\,\beta'!\,(A-\alpha-\alpha')!\,(N-\beta-\beta')!\,(A+N)!}$$

Probabilitatea ca să avem $x = \alpha'$, în primele $\alpha' + \beta'$ extracții și $y = \alpha$ în următoarele $\alpha + \beta$ extracții, este

$$\frac{C_A^{\alpha'} \cdot C^{\beta'}}{C_{A+N}^{\alpha'+\beta'}} \cdot \frac{C_{A-\alpha'}^{\alpha} \cdot C_{N-\beta'}^{\beta}}{C_{A+N-\alpha'-\beta'}^{\alpha+\beta}} = \frac{A!\cdot N!\,(\alpha+\beta)!\,(A+N-\alpha-\beta-\alpha'-\beta)\,(\alpha'+\beta')!}{\alpha!\,\beta!\,\alpha'!\,\beta'!\,(A-\alpha-\alpha')!\,(N-\beta-\beta)!\,(A+N)!}.$$

După cum se vede, cele două probabilități sînt egale.

Elementele de mai sus ne sînt insuficiente pentru a putea determina formulele de calcul ale momentelor inițiale și centrate de diferite ordine.

a) *Momente inițiale.* Vom introduce notația

$$M_{rs} = M(x^r \cdot y^s)$$

pentru valoarea medie a produsului $x^r \cdot y^s$ care este mai comodă decît notația $\overline{x^r \cdot y^s}$.

Păstrînd principiul de calcul a valorii medii a unui produs de două variabile dintr-un sistem, potrivit formulei (IV, 3.1) vom avea

$$\mathrm{M}_{rs} = \sum_{k=1}^{m} \sum_{j=1}^{n} p_{kj}\, a_k^r\, b_j^s. \qquad \text{(XI, 2.7)}$$

După cum se vede, respectînd nomenclatura adoptată pentru distribuții de o singură caracteristică, vom putea spune că M_{rs} este un *moment inițial de ordinul* $r + s$ al distribuției cu două caracteristici sau momentul de ordinul (r, s). Pentru momentul M_{00} vom avea evident $M_{00} = 1$.

Momentele inițiale de ordinul întîi sînt:

$$\left.\begin{aligned} M_{10} &= \sum_{k=1}^{m} \sum_{j=1}^{n} p_{kj}\, a_k = \sum_{k=1}^{m} a_k \sum_{j=1}^{n} p_{kj} = \sum_{k=1}^{m} p_{k\cdot}\, a_k \\ M_{01} &= \sum_{k=1}^{m} \sum_{j=1}^{n} p_{kj}\, b_j = \sum_{j=1}^{n} b_j \sum_{k=1}^{m} p_{kj} = \sum_{j=1}^{n} p_{\cdot j}\, b_j \end{aligned}\right\} \qquad \text{(XI, 2.8)}$$

Pentru simplificare putem folosi și notația:

$$M_{10} = m_1 \qquad M_{01} = m_2 .$$

Punctul din plan avînd coordonatele (m_1, m_2) poartă numele de *centru de greutate* al distribuției de două caracteristici. El joacă același rol ca și media aritmetică pentru distribuțiile de o singură caracteristică.

Scanned with OKEN Scanner

Reprezentarea grafică a centrului de greutate este arătată în figura 25.

Punctul C de coordonate (M_{01}, M_{10}) reprezintă centrul de greutate al unei distribuții oarecare de două caracteristici x și y. Mulțimea de puncte, dispersate în plan reprezintă celelalte cazuri de perechi de valori de coordonate $x = a_k$, $y = b_j$ care se concentrează mai mult sau mai puțin în jurul centrului de greutate.

Momentele inițiale de ordinul doi sînt:

$$\left.\begin{aligned} M_{20} &= \sum_{k=1}^{m}\sum_{j=1}^{n} p_{kj}\, a_k^2 = \sum_{k=1}^{m} p_{k\cdot}\, a_k^2, \\ M_{11} &= \sum_{k=1}^{m}\sum_{j=1}^{n} p_{kj}\, a_k\, b_j, \\ M_{02} &= \sum_{k=1}^{m}\sum_{j=1}^{n} p_{kj}\, b_j^2 = \sum_{j=1}^{n} p_{\cdot j}\, b_j^2. \end{aligned}\right\} \quad \text{(XI, 2.9)}$$

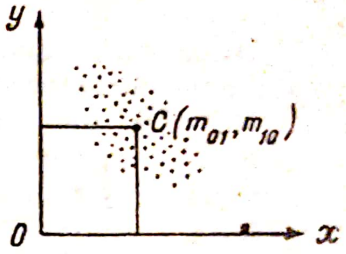


Fig. 25

Putem remarca un caz particular, cînd x și y sînt independente între ele, adică atunci cînd este îndeplinită relația (XI, 2.5) este îndeplinită și relația (XI, 2.6).

Într-adevăr, avem

$$\begin{aligned} M_{11} &= \sum_{k=1}^{m}\sum_{j=1}^{n} p_{kj}\, a_k\, b_j = \sum_{k=1}^{m}\sum_{j=1}^{n} p_{k\cdot} p_{\cdot j}\, a_k\, b_j = \\ &= \sum_{k=1}^{m} p_{k\cdot}\, a_k \sum_{j=1}^{n} p_{\cdot j}\, b_j = M_{10}\, M_{01} = m_1 \cdot m_2 . \end{aligned} \quad \text{(XI,2.10)}$$

Se vede că egalitatea (XI, 2.10) este identică cu egalitatea (XI, 2.6) ținînd seamă de notațiile adoptate.

b) *Momente centrate.* După cum știm momentele centrate sînt definite, în general, prin egalitatea

$$M_{rs} = M\,(x - m_1)^r (y - m_2)^s . \quad \text{(XI, 2.11)}$$

sau

$$M_{rs} = \sum_{k=1}^{m}\sum_{j=1}^{n} p_{kj}\,(a_k - m_1)^r (b_j - m_2)^s .$$

Din această formulă generală rezultă imediat primele momente centrate.

Momente centrate de ordinul întîi:

$$\left.\begin{aligned} \mu_{10} &= \sum_{k=1}^{m}\sum_{j=1}^{n} p_{kj}(a_k - m_1) = 0, \\ \mu_{01} &= \sum_{k=1}^{m}\sum_{j=1}^{n} p_{kj}(b_j - m_2) = 0. \end{aligned}\right\} \quad \text{(XI, 2.12)}$$

Momente centrate de ordinul doi:

$$\left.\begin{aligned}\mu_{20} &= \sum_{k=1}^{m}\sum_{j=1}^{n} p_{kj}(a_k - m_1)^2 = \sum_{k=1}^{m} p_{k\cdot}(a_k - m_1)^2\\ \mu_{11} &= \sum_{k=1}^{m}\sum_{j=1}^{n} p_{kj}(a_k - m_1)(b_j - m_2),\\ \mu_{02} &= \sum_{k=1}^{m}\sum_{j=1}^{n} p_{kj}(b_j - m_2)^2 = \sum_{j=1}^{n} p_{\cdot j}(b_j - m_2)^2.\end{aligned}\right\} \quad \text{(XI, 2.13)}$$

După cum știm, momentele centrate de ordinul doi reprezintă dispersia unei caracteristici variabile.

Din relațiile (XI, 2.13) se vede că:

$m\mu_{20} = \sigma_1^2$ reprezintă dispersia caracteristicii în x;

$m\mu_{02} = \sigma_2^2$ reprezintă dispersia caracteristicii în y;

$m\mu_{11} =$ reprezintă covarianța componentelor x și y. În cazul cînd x și y sînt independente, $\mu_{11} = 0$.

Pentru o distribuție de două caracteristici, funcția caracteristică $\varphi(t, u)$ este definită prin relația;

$$\varphi(t, u) = M(e^{(tx+uy)i}),$$

în care t și u sînt variabile reale,

$$\varphi(t, u) = \sum_{k=1}^{m}\sum_{j=1}^{n} p_{kj}\, e^{(a_k t + b_j u)i} \quad \text{(XI, 2.14)}$$

3. Valori medii condiționate. Funcție de regresie

Vom nota prin $M(x/y = b_j)$, valoarea medie a lui x condiționată de egalitatea $y = b_j$, și prin $M(y/x = a_i)$, valoarea medie a lui y condiționată de $x = a_i$.

Ținînd seama de relațiile (XI,2.3) și (XI,2.4) rezultă imediat

$$\left.\begin{aligned}M(x/y = b_j) &= \sum_{i=1}^{m} a_i\, p(x = a_i/y = b_j) = \frac{\sum_{i=1}^{m} p_{ij} a_i}{p_{\cdot j}},\\ M(y/x = a_i) &= \sum_{j=1}^{n} b_j\, p(y = b_j/x = a_i) = \frac{\sum_{j=1}^{n} p_{ij} b_j}{p_{i\cdot}}.\end{aligned}\right\} \quad \text{(XI, 3,1)}$$

După cum se vede, $M(x/y = b_j)$ este o funcție de valorile b_j.

Pentru fiecare valoare a lui b_j există o valoare corespunzătoare

$$\alpha_j = M(x/y = b_j), \qquad (j = 1, 2, \ldots, n).$$

Această funcție poartă numele de *funcție de regresie a mediei lui x asupra lui y.*

Punctele de coordonate (α_j, b_j) $(j = 1, 2, \ldots, n)$ ne dau, prin unirea lor cu segmente de linii drepte, *funcția de regresie a lui x asupra lui y*

În mod analog $M(y/x = a_i)$ este o funcție de a_i. Pentru fiecare valoare a lui a_i există o valoare $\beta_i = M(y/x = a_i)$, $(i = 1, 2, \ldots, m)$, care ne dă *funcția de regresie a mediei lui y asupra lui x.*

În loc să considerăm media condiționată a valorilor lui y putem să folosim și alte medii.

Dacă notăm prin $M_e(x/y = b_j)$ mediana valorilor lui x condiționată de egalitatea $y = b_j$, mărimea $M_e(x/y = b_j)$ este o funcție de b_j și care poartă numele de *funcție de regresie a medianei lui x asupra lui y.*

În mod similar, $M_e(y/x = a_i)$ reprezintă mediana valorilor y, condiționată de egalitatea $x = a_i$, purtînd numele de *funcție de regresie a medianei lui y asupra lui x.*

Regresia mediei și regresia medianei se bucură de anumite proprietăți extremale, specifice acestei valori medii.

Să considerăm astfel expresia

$$M_2(x - \alpha/y = b_j)$$

care reprezintă momentul de ordinul al doilea al lui $x - \alpha$, condiționat de $y = b_j$. Știm că acest moment ia cea mai mică valoare pentru

$$\alpha_j = M(x/y = b_j).$$

Prin urmare, dacă ne propunem de a determina o funcție $\alpha = \varphi(b)$ astfel ca expresia

$$M_2(x - \varphi(b_j)/y = b_j)$$

să devină minimă, $\varphi(b)$ este funcția de regresie a mediei lui x asupra lui y.

În mod analog, momentul condiționat de ordinul al doilea

$$M_2(y - \psi(a_i)/x = a_i)$$

devine minim pentru

$$\psi(a_i) = M_2(y/x = a_i).$$

Aceleași considerații putem face și pentru funcția de regresie a medianei. Am văzut că mediana face minimă suma abaterilor, în valoare absolută, ale unei distribuții.

Prin urmare, dacă vom considera expresia

$$M(|x - \varphi(b_j)|/y = b_j),$$

minimul ei este atins pentru

$$\varphi(b_j) = M_e(x/y = b_j).$$

În mod analog, funcția

$$M(|y - \psi(a_i)|/x = a_i)$$

devine minimă pentru

$$\psi(a_i) = M_e(y/x = a_i).$$

4. Analiza legăturii statistice dintre caracteristicile distribuției Metoda corelației

A. *Rolul analizei calitative.* Din cele expuse mai sus am constatat că între cele două caracteristici ale distribuției, considerate ca variabile aleatoare, pot exista legături de dependență sau dimpotrivă, asemenea

legături să nu existe, caracteristicile variabile fiind independente. Criteriul independenței pe plan teoretic îl constituie relația (XI, 2. 5)

În studiul vieții social-economice, ne întîlnim — în majoritatea cazurilor — cu fenomene interdependente, condiționate reciproc, pe care statistica prin metodele ei proprii, este chemată să le caracterizeze, estimînd forma legăturii de interdependență și măsurînd intensitatea ei. Tocmai în acest domeniu, pornind de la unitatea dialectică a laturilor calitative și cantitative ale fenomenelor social-economice, statistica își vădește caracterul ei de știință.

Între fenomene pot exista legături variate. Caracterul acestor legături, natura lor specifică, nu poate fi însă stabilită decît printr-o analiză teoretică, calitativă, multilaterală, iar concluziile ce se vor trage pe baza unei asemenea analize vor avea un rol hotărîtor în relevarea legăturilor obiective și necesare care există între fenomenele studiate ca și în determinarea formei matematice în care ea poate fi exprimată.

Vom accentua și aici că analiza teoretică, calitativă în domeniul fenomenelor vieții social-economice se fundamentează pe tezele materialismului istoric și al economiei politice marxiste. Aceste teze ne pot dezvălui faptul că, în unele cazuri, între fenomenele cercetate există o legătură cauzală, iar în altele o legătură de interdependență de condiționare reciprocă — cînd fiecare dintre fenomene este în același timp și cauză și efect; de asemenea, faptul că variația paralelă a două fenomene nu se datorește legăturii dintre ele ci influenței unui al treilea fenomen sau faptul că între două sau mai multe fenomene există o legătură funcțională al cărui caracter, după cum știm, este unitar. În sfîrșit, analiza calitativă ne poate dezvălui existența legăturilor statistice dintre caracteristicile studiate. Așa după cum s-a arătat de la începutul acestui capitol, legăturile statistice indică o corespondență deosebită — de la o valoare a uneia dintre caracteristici la o distribuție întreagă de valori ale alteia. Legăturile statistice sînt expresia legităților statistice care acționează în viața social-economică, determinate, la rîndul lor, de un ansamblu întreg de condiții în care se manifestă fenomenul de masă.

Legitatea statistică se manifestă și poate fi descoperită numai într-un mare număr de cazuri observate cînd dispar abaterile întîmplătoare într-o direcție sau alta, compensîndu-se reciproc, și cînd iese la iveală tendința generală, căreia i se supun fenomenele individuale.

Referindu-ne la cazul concret al distribuțiilor de două caracteristici, analiza calitativă ne poate indica forma concretă a ecuației de corelație dintre caracteristicile studiate ca de pildă între productivitatea muncii și prețul de cost, între productivitatea muncii și gradul de înzestrare tehnică a forței de muncă, între salariul și vechimea în cîmpul muncii a muncitorilor, între gradul de îndeplinire a planului producției globale și gradul de utilizare a timpului de lucru etc.

În cazul corelației cauzale, una din caracteristici este independentă în raport cu cealaltă și se notează de obicei cu x, iar cealaltă este dependentă de prima și se notează cu y. De obicei caracteristica independentă se numește factorială iar cea dependentă, rezultativă.

Scanned with OKEN Scanner

Desigur, în procesul măsurării statistice a legăturii dintre cele două caracteristici pot fi aduse unele corectări privind forma de exprimare a acestei legături, însă asemenea corectări nu pot afecta în nici un caz concluzia de fond care s-a desprins în mod dialectic din analiza naturii calitative specifice fenomenului cercetat.

B. *Metoda corelației.* Ca și metoda grupărilor, a mediilor etc., metoda corelației este specifică statisticii — elaborată pentru studiul legăturilor dintre caracteristicile din sistem. Cu ajutorul metodei corelației se exprimă numeric direcția și intensitatea legăturii, scoțîndu-se în evidență legitatea care se manifestă în cadrul ei.

Două probleme esențiale se pun cu acest prilej.

În primul rînd, se pune problema de a se releva pe baza datelor de masă ale observării, cum anume se schimbă caracteristica rezultativă — funcția — prin efectul schimbării uneia din caracteristicile independente, dacă influența celorlalte variabile independente este înlăturată. Aceasta nu înseamnă, desigur, că celelalte caracteristici independente — secundare în raport cu obiectul principal al cercetării — își încetează realmente influența; este vorba de aplicarea metodei științifice de abstractizare care-și găsește o largă aplicație și în studiul corelației statistice. În rezolvarea acestor probleme, influența denaturantă a factorilor străini de corelația care constituie obiectul cercetărilor noastre se manifestă, după caz, cu o intensitate mai mare sau mai mică.

Este necesar, de aceea, în al doilea rînd, să măsurăm și această influență pentru că, indirect, ea ne va oglindi tocmai puterea cu care își croiește drum legitatea care ne interesează.

Este evident că studiul corelației se complică, o dată cu creșterea numărului de caracteristici din sistem. În continuare ne vom opri, mai întîi asupra corelației ce se manifestă în cadrul distribuțiilor cu două caracteristici, adică asupra așa-numitei *corelații simple*, care are multiple aplicații în practică.

Corelația simplă. Fiind vorba de variația a două caracteristici, materialul statistic de bază trebuie sistematizat și grupat în așa fel ca să iasă în evidență tocmai caracteristicile ce ne interesează. După cum am văzut, acest lucru se realizează prin întocmirea tabelelor statistice cu două intrări ca și prin elaborarea graficelor. O formă de sistematizare grafică a materialului este și aceea a reprezentării *cîmpului de corelație* prin determinarea în plan a mulțimii de puncte de coordonate x_i, y_i.

Cu ajutorul grupării intervalelor de variație a funcției statistice și apoi a calculării mediilor $\bar{y}_i$, pe cîmpul de corelație se poate întocmi o linie frîntă care va reprezenta linia statistică de regresie.

Un interesant caz de corelație simplă se vede în tabelul 36 *), în care sînt grupate datele caracterizînd situația în 214 întreprinderi industriale din regiunea Kuibîșev, după două caracteristici variabile: procentul de îndeplinire a planului de om — zile muncă (x) și procentul de îndeplinire a planului producției globale (y).

*) Acest exemplu este luat după I. I. Lukomski din lucrarea sa *Teoria korreliații i primenenie k analizu proizvodstva*. Moskva, 1958, prima lui sursă este lucrarea lui K. Fastov, publicată în Ucenîe zapisku planovogo instituta, vol. 2, Kuibîșev, 1940.

Tabelul 36

Pentru 214 întreprinderi industriale din regiunea Kuibișev (1935). Caracteristici: producția globală și numărul de om — zile (% față de plan)

y \ x	Grupe de intervale ale procentului de îndeplinire a planului de om — zile muncă (x)												
	Sub 55**)	55—65	65—75	75—85	85—95	95—105	105—115	115—125	125—135	135—145	145—155	Peste 155*)	Total
Grupe de intervale ale procentului de îndeplinire a planului producției globale (y)													
Peste 155*)	—	—	—	—	1	1	2	1	—	—	1	4	10
145—155	—	—	—	—	—	1	—	—	—	—	2	3	6
135—145	—	—	—	—	—	—	—	4	1	—	1	2	8
125—135	—	—	—	—	4	4	3	5	4	—	—	—	20
115—125	—	—	—	1	—	8	5	6	2	1	1	2	26
105—115	—	1	—	3	9	15	6	4	1	1	—	2	42
95—105	—	—	1	6	6	20	8	2	—	1	—	—	44
85— 95	—	—	1	6	5	9	5	2	—	1	—	1	30
75— 85	—	—	1	1	5	3	2	—	—	—	—	—	12
65— 75	3	3	1	1	—	—	—	—	—	—	—	—	8
55— 65	1	2	—	1	1	—	—	—	—	—	—	—	5
Sub 55**)	2	1	—	—	—	—	—	—	—	—	—	—	3
Total:	6	7	4	19	31	61	31	24	8	4	5	14	214

*) Acest interval se consideră închis între limitele 155—165.
**) Acest interval se consideră închis între limitele 45—55.

Tabelul 37

Valorile pereche ale lungimii de rupere și ai gradului de măcinare S.R. pt. dif. intervale de grad de măcinare

Nr. crt. al experienței / Limitele intervalului de variație al gradului de măcinare (x)	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
28—32	1 625 29	1 480 31	1 425 30	1 580 31	1 610 30	1 670 30	1 540 32	1 530 29	1 625 31	1 560 30	1 590 30	1 755 31	1 535 30	1 510 28	1 470 30
33—37	1 560 36	1 720 35	1 565 35	1 530 34	1 615 37	1 600 35	1 535 36	1 455 35	1 625 34	1 710 36	1 520 35	1 550 35	1 535 33	1 630 35	1 505 35
38—42	1 620 39	1 800 40	1 515 39	1 600 38	1 610 41	1 655 40	1 580 38	1 630 40	1 580 39	1 780 42	1 570 40	1 605 40	1 660 41	1 550 40	1 635 39
42—47	1 605 45	1 695 44	1 720 46	1 780 45	1 820 47	1 755 45	1 655 45	1 740 44	1 680 45	1 790 46	1 665 44	1 700 45	1 715 46	1 725 45	1 710 45
48—52	1 785 51	1 990 50	1 790 49	1 945 50	1 825 48	1 940 51	1 870 50	1 835 50	1 780 50	1 865 49	1 910 49	1 805 50	1 665 52	1 860 51	1 900 50
53—57	1 945 55	1 970 55	2 015 55	1 990 56	2 050 54	2 135 57	2 135 55	2 065 55	1 900 54	1 970 53	1 980 55	1 985 54	2 050 56	2 160 55	2 040 55
58—62	2 160 62	2 510 60	2 190 59	2 215 59	2 150 61	2 365 60	2 335 60	2 180 58	2 195 61	2 180 60	2 275 59	2 250 61	2 145 60	2 370 61	2 110 60
63—67	2 560 64	2 280 65	2 650 66	2 340 65	2 390 65	2 485 64	2 475 65	2 355 63	2 520 65	2 600 65	2 675 65	2 440 64	2 590 66	2 735 65	2 405 64
68—72	2 720 71	3 030 70	2 665 69	2 810 71	2 890 70	2 595 72	2 755 70	2 940 71	2 800 69	2 625 70	2 865 69	2 530 70	3 000 71	2 840 70	2 935 69
73—77	3 280 76	3 035 74	3 340 75	3 265 75	2 990 75	3 150 74	3 290 76	3 200 75	3 160 75	3 215 74	3 160 76	3 110 75	3 060 75	2 920 75	2 955 73

Scanned with OKEN Scanner

După cum se vede din acest tabel (primele coloane), întreprinderile care au un nivel scăzut al gradului de îndeplinire a planului de om-ore muncă (x) se caracterizează de asemenea și printr-un nivel scăzut al gradului de îndeplinire a planului producției globale (y). Pe coloanele următoare se observă că la serii de întreprinderi la care cresc valorile procentului de îndeplinire a planului de om-ore muncă, cresc și valorile procentului de îndeplinirea planului producției globale.

Se poate spune deci că între cele două caracteristici ilustrate în tabel, există o *corelație directă* (la alte distribuții poate fi observat un fenomen invers — creșterii valorilor caracteristice factoriale îi corespund descreșteri ale valorilor caracteristicii rezultative, în care caz se spune că există o *corelație inversă*).

Ceea ce caracterizează, deci, o asemenea corelație este:

a) dependența directă în același sens a valorilor funcției (y), de cele ale caracteristicii independente (x);

b) fiecărei valori a caracteristicii independente îi corespunde o serie de distribuții a funcției;

c) odată cu creșterea valorilor caracteristicii independente, seriile de distribuție ale funcției se deplasează, schimbîndu-și poziția.

În legătură cu acest ultim aspect al problemei trebuie să arătăm că nu avem nici o corelație dacă o dată cu creșterea valorilor caracteristicii factoriale (x) seriile de distribuții ale funcției (y) nu-și schimbă poziția sau o schimbă cu totul întîmplător.

Un alt interesant exemplu de corelație se găsește și într-o lucrare publicată de N. Rancu și L. Tövissi*) din activitatea industriei hîrtiei.

În tabelul 37 s-au trecut rezultatele a mai multor experiențe privind lungimea de rupere și gradul de mărime S. R. (Schoper-Riegler) pentru hîrtia de ziar.

În baza acestui tabel s-a întocmit tabelul de corelație 38 unde x reprezintă lungimea de rupere, iar y gradul de măcinare.

Tabelul 38

y \ x	30	35	40	45	50	55	60	65	70	75	Total
3 200—3 400	—	—	—	—	—	—	—	—	—	5	5
3 000—3 200	—	—	—	—	—	—	—	—	1	7	8
2 800—3 000	—	—	—	—	—	—	—	—	7	3	10
2 600—2 800	—	—	—	—	—	—	—	3	5	—	8
2 400—2 600	—	—	—	—	—	—	1	8	2	—	11
2 200—2 400	—	—	—	—	—	—	6	4	—	—	10
2 000—2 200	—	—	—	—	—	8	8	—	—	—	16
1 800—2 000	—	—	—	1	11	7	—	—	—	—	19
1 600—1 800	5	5	9	14	4	—	—	—	—	—	37
1 400—1 600	10	10	6	—	—	—	—	—	—	—	26
											150

*) N. Rancu și I. Tövissi, *Folosirea metodei corelației la studiul, dirijarea și îmbunătățirea proceselor tehnologice din industria hîrtiei*. Celuloză și hîrtie, *5* (1958).

În figura 26 s-a întocmit un grafic de corelație, care arată modul de răspîndire al punctelor din tabelul 38 în plan. Se vede ușor că distribuția valorilor se grupează în jurul unei parabole. Curba de regresie în cazul

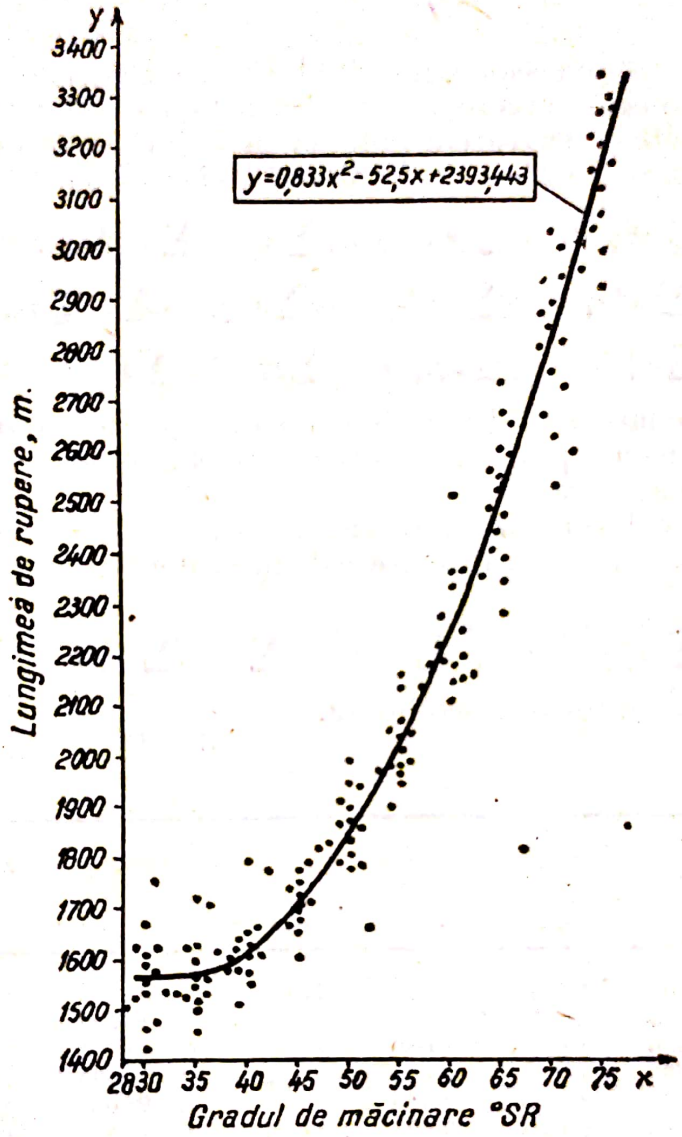


Fig. 26

de față pare să fie o parabolă și de aceea se încearcă o ajustare prin aplicarea formulei

$$y = a_0x^2 + a_1x + a_2.$$

Va trebui deci să se determine coeficienții a_0, a_1, a_2.
Punînd

$$\bar{y}_i = a_0\, x_i^2 + a_1x_i + a_2$$

și condiția pătratelor minime,

$$\sum(y_i - \bar{y}_i)^2 = \text{minim},$$

ajungem la următoarea problemă : să se determine parametrii a_0, a_1, a_2 astfel ca expresia

$$\sum_i (y_i - a_0 x_i^2 - a_1 x_i - a_2)^2 \qquad \text{(XI, 4.1)}$$

să fie minimă.

Se știe că o sumă este minimă atunci cînd derivatele în raport cu necunoscutele sînt egale cu zero. În expresia de mai sus sînt trei necunoscute a_0, a_1 și a_2. Calculînd derivatele funcției în raport cu aceste necunoscute și egalîndu-le cu zero, se obțin următoarele trei ecuații cu trei necunoscute :

$$\begin{aligned} a_0 \sum x_i^2 f_i + a_1 \sum x_i f_i + a_2 \sum f_i &= \sum \bar{y}_i f_i, \\ a_0 \sum x_i^3 f_i + a_1 \sum x_i^2 f_i + a_2 \sum x_i f_i &= \sum x_i \bar{y}_i f_i, \\ a_0 \sum x_i^4 f_i + a_1 \sum x_i^3 f_i + a_2 \sum x_i^2 f_i &= \sum x_i^2 \bar{y}_i f_i. \end{aligned} \qquad \text{(XI, 4.2)}$$

În aceste ecuații x_i sînt variabile obținute pentru gradul de măcinare, iar f indică numărul probelor pentru care s-a obținut aceeași valoare a gradului de măcinare.

Ecuațiile pe baza cărora se calculează coeficienții a_0, a_1, a_2 se pot scrie direct, fără a fi necesar să se calculeze derivatele parțiale. Valorile sumelor

$$\sum x_i f_i, \sum x_i^2 f_i, \sum x_i^3 f_i, \sum x_i^4 f_i \text{ și } \sum \bar{y} f_i, \sum x_i \bar{y}_i f_i, \sum x_i^2 \bar{y}_i f_i$$

se vor calcula cu ajutorul tabelului 39.

Tabelul 39

x	x_i	$\sum y_i$	$\bar{y}_i$	x_i^2	x_i^3	x_i^4	$x_i \bar{y}_i$	$x_i^2 \bar{y}_i$
1	2	3	4	5	6	7	8	9
28—32	30	23 505	1 567	900	27 000	810 000	47 010	1 410 300
33—37	35	23 655	1 577	1 225	42 875	1 500 625	55 195	1 931 825
38—42	40	24 390	1 626	1 600	64 000	2 560 000	65 040	2 601 600
43—47	45	25 755	1 717	2 025	91 125	4 100 625	77 265	3 476 925
48—52	50	21 765	1 851	2 500	125 000	6 250 000	92 550	4 627 500
53—57	55	30 390	2 026	3 025	166 375	9 150 625	111 430	6 128 650
58—62	60	33 730	2 242	3 600	216 000	12 960 000	134 520	8 071 200
63—67	65	37 500	2 500	4 225	274 625	17 860 625	162 500	10 562 500
68—72	70	42 000	2 800	4 900	343 000	24 010 000	196 000	13 720 000
73—77	75	47 130	3 142	5 625	421 875	31 640 625	235 650	17 673 750
Total	525		21 048	29 625	1 771 875	110 833 125	1 177 160	70 204 250

Pentru ușurință, în loc să se folosească toate valorile lui x obținute într-un interval, se ia numai centrul intervalului. Centrele intervalelor sînt date în coloana 2 a tabelului. În acest caz, trebuie să se calculeze media lungimilor de rupere $\bar{y}_i$. Aceasta înseamnă că în locul celor 15 puncte corespunzătoare unui interval al gradului de măcinare va fi un singur punct a cărui poziție este determinată de o pereche de valori ; centrul inter-

Scanned with OKEN Scanner

valului x_i și mediei lungimii de rupere ($\bar{y}_i$). Valorile lui $\bar{y}_i$ sînt date în coloana 4 a tabelului și se obțin împărțind suma valorilor lui $\bar{y}_i$ corespunzătoare unui interval dat (datele coloanei 3 din tabel) la numărul valorilor lui x cuprinse în acest interval f_i. Se observă că fiecare interval conține 15 valori ale lui x, deci f_i egal cu 15 este o cantitate constantă, fapt care permite să se împartă cele trei ecuații ale sistemului (XI, 4.2) prin 15 și să se obțină sistemul

$$\begin{aligned} a_0 \sum x_i^2 + a_1 \sum x_i + 10\, a_2 &= \sum \bar{y}_i, \\ a_0 \sum x_i^3 + a_1 \sum x_i^2 + a_2 \sum x_i &= \sum x_i \bar{y}_i, \\ a_0 \sum x_i^4 + a_1 \sum x_i^3 + a_2 \sum x_i^2 &= \sum x_i^2 \bar{y}_i \end{aligned} \qquad \text{(XI, 4.3)}$$

$\sum x_i$, $\sum x_i^2$, $\sum x_i^3$, $\sum x_i^4$, $\sum \bar{y}_i$, $\sum x_i \bar{y}_i$, $\sum x_i^2 \bar{y}_i$ se obțin făcînd totalul coloanelor 2, 5, 6, 7, 4, 8 și 9 ale tabelului. Făcînd înlocuirile în sistemul de ecuații (XI, 4.3), se obține

$$\begin{aligned} 29\,625\, a_0 + 525\, a_1 + 10\, a_2 &= 21\,048, \\ 1\,771\,875\, a_0 + 29\,625\, a_1 + 525\, a_2 &= 1\,177\,160, \\ 110\,833\,125\, a_0 + 1\,771\,875\, a_1 + 29\,625\, a_2 &= 70\,204\,250. \end{aligned} \qquad \text{(XI, 4.4)}$$

Soluția acestui sistem este:

$$a_0 = \frac{\begin{vmatrix} 21\,048 & 525 & 10 \\ 1\,171\,160 & 29\,625 & 525 \\ 70\,204\,250 & 1\,771\,875 & 29\,625 \end{vmatrix}}{\begin{vmatrix} 29\,625 & 525 & 10 \\ 1\,771\,875 & 29\,625 & 525 \\ 110\,833\,125 & 1\,771\,875 & 29\,625 \end{vmatrix}} = 0{,}833,$$

$$a_1 = \frac{\begin{vmatrix} 29\,625 & 21\,048 & 10 \\ 1\,771\,875 & 1\,177\,160 & 525 \\ 110\,833\,125 & 70\,204\,250 & 29\,625 \end{vmatrix}}{\begin{vmatrix} 29\,625 & 525 & 10 \\ 1\,771\,875 & 29\,625 & 525 \\ 110\,833\,125 & 1\,771\,875 & 29\,625 \end{vmatrix}} = 52{,}5,$$

$$a_2 = \frac{\begin{vmatrix} 29\,625 & 525 & 21\,048 \\ 1\,771\,875 & 29\,625 & 1\,177\,160 \\ 110\,833\,125 & 1\,771\,875 & 70\,204\,250 \end{vmatrix}}{\begin{vmatrix} 29\,625 & 525 & 10 \\ 1\,771\,875 & 29\,626 & 525 \\ 110\,833\,125 & 1\,771\,875 & 29\,625 \end{vmatrix}} = 2\,393{,}443.$$

Înlocuind valorile lui a_0, a_1 și a_2 se obține

$$y_i = 0{,}833\, x_i^2 - 52{,}5\, x_i + 2\,393{,}443.$$

5. Drepte de regresie

În același mod în care am determinat în cazul special de mai sus o curbă parabolică de regresie, putem să determinăm și ecuațiile unor drepte de regresie.

Vom trata de data aceasta problema în cazul general.

Să considerăm o distribuție teoretică discontinuă de două variabile, ale cărei variabile x și y pot lua, respectiv, valorile a_i, $(i = 1, 2, \ldots, m)$, și b_j, $(j = 1, 2, \ldots, n)$, cu probabilitățile

$$P(x = a_i,\ y = b_j) = p_{ij}.$$

Să determinăm parametrii arbitrari α și β astfel ca expresia

$$M(y - \alpha x - \beta)^2 \qquad \text{(XI, 5.1)}$$

să fie minimă.

Expresia $M(y - \alpha x - \beta)^2$ se mai poate scrie

$$\sum_{i=1}^{m}\sum_{j=1}^{n} p_{ij}\,(b_j - \alpha a_i - \beta)^2. \qquad \text{(XI, 5.2)}$$

Atunci problema se reduce la determinarea minimului expresiile (XI, 5.2). Trebuie, pentru aceasta, să anulăm derivatele parțiale de ordinul întîi în raport cu α și β ale acestei expresii, adică

$$\begin{aligned} &\sum_{i=1}^{m}\sum_{j=1}^{n} p_{ij}\,a_i\,(b_i - \alpha a_i - \beta) = 0, \\ &\sum_{i=1}^{m}\sum_{j=1}^{n} p_{ij}\,(b_j - \alpha a_i - \beta) = 0. \end{aligned} \qquad \text{(XI, 5.3)}$$

Relațiile (XI, 5.3) se mai pot scrie

$$\begin{aligned} M_{11} - \alpha\, M_{20} - \beta m_1 &= 0 \\ m_2 - \alpha\, m_1 - \beta &= 0 \end{aligned}$$

sau

$$\begin{aligned} \alpha\, M_{20} + \beta m_1 &= M_{11} \\ \alpha\, m_1 + \beta &= m_2. \end{aligned} \qquad \text{(XI, 5.4)}$$

Rezolvînd acest sistem obținem

$$\alpha = \frac{M_{11} - m_1 m_2}{M_{20} - m_1^2}, \qquad \beta = \frac{m_2\, M_{20} - m_1\, M_{11}}{M_{20} - m_1^2}.$$

Dacă notăm

$$\sigma_1^2 = M_{20} - m_1^2;\ \sigma_2^2 = M_{02} - m_2^2.$$

$$\rho = \frac{M_{11} - m_{11}}{\sigma_1\,\sigma_2}$$

soluțiile de mai sus devin

$$\alpha = \frac{\rho\sigma_i}{\sigma_1}, \qquad \beta = \frac{m_2\,\sigma_1 - m_1\,\rho\,\sigma_2}{\sigma_1}.$$

Am obținut, astfel, dreapta de regresie

$$y - \frac{\rho\,\sigma_2}{\sigma_1}x - \frac{m_2\,\sigma_1 - m_1\,\rho\,\sigma_2}{\sigma_1} = 0$$

sau

$$y - m_2 = \frac{\rho\,\sigma_2}{\sigma_1}(x - m_1). \qquad \text{(XI, 5.5)}$$

În mod analog, dacă ne-am fi propus să determinăm parametrii arbitrari α și β în așa fel ca expresia

$$M\,(x - \alpha y - \beta)^2$$

să fie minimă, am fi obținut dreapta de regresie

$$y - m_2 = \frac{\rho\,\sigma_1}{\sigma_2}(x - m_1). \qquad \text{(XI, 5.6)}$$

Cele două drepte, definite prin ecuațiile (XI, 5.5) și (XI, 5.6) sînt dreptele de regresie clasic care, după cum se vede, trec prin centrul de greutate (m_1, m_2) al distribuției.

În general, aceste drepte de regresie sînt diferite, afară de cazul cînd

$$\frac{\rho\,\sigma_1}{\sigma_2} = \frac{\rho\,\sigma_2}{\sigma_1}$$

sau

$$\sigma_1 = \sigma_2.$$

Prin urmare, cele două drepte de regresie vor coincide, dacă dispersiile σ_1 și σ_2 sînt egale.

Cunoscînd valorile parametrilor α și β, pentru care expresia (XI, 5.1) este minimă, vom putea să calculăm acest minim; astfel, minimul expresiei (XI, 5.1) este

$$M\left[y - m_2 - \frac{\rho\,\sigma_2}{\sigma_1}(x - m_1)\right]^2 = M\,(y - m_2)^2 - 2\,\frac{\rho\,\sigma_2}{\sigma_1}M\,(x - m_1)(y - m_2) +$$

$$+ \frac{\rho^2\,\sigma_2^2}{\sigma_1^2}M\,(x - m_1)^2 = \sigma_2^2 - \frac{2\,\rho\,\sigma_2}{\sigma_1}\rho\,\sigma_1\,\sigma_2 + \frac{\rho^2\,\sigma_2^2}{\sigma_1^2}\sigma_1^2 = \sigma_2^2\,(1 - \rho^2),\ (\rho \leqslant 1).$$

Minimul expresiei $M(x - \alpha\,y - \beta)^2$ este egal cu

$$M\left[y - m_2 - \frac{\rho\,\sigma_1}{\sigma_2}(x - m_1)\right]^2 = \sigma_1^2\,(1 - \rho^2),\ (\rho \leqslant 1).$$

6. Coeficientul de corelație

În practica statistică se introduc anumiți coeficienți, care pot să dea o indicație asupra gradului de dependență a celor două caracteristici dintr-un tabel de corelație. Un asemenea indicator este și coeficientul de corelație ρ dat de expresia

$$\rho = \frac{M\,[(x - m_1)(y - m_2)]}{\sqrt{M(x - m_1)^2}\,\sqrt{M(y - m_2)^2}}. \qquad \text{(XI, 6.1)}$$

cunoscut în statistică sub numele de coeficient de corelație. El se bucură de mai multe proprietăți:

1°. Coeficientul de corelație ρ satisface inegalitățile

$$-1 \leqslant \rho \leqslant 1.$$

Într-adevăr, să considerăm următoarea formă pătratică pozitiv definită

$$\sum_{i=1}^{m}\sum_{j=1}^{n} p_{ij}\,[\lambda(a_i - m_1) + \mu(b_j - m_2)]^2 = \lambda^2 \sum_{i=1}^{m}\sum_{j=1}^{n} p_{ij}(a_i - m_1)^2 +$$

$$+ 2\lambda\mu \sum_{i=1}^{m}\sum_{j=1}^{n} p_{ij}(a_i - m_1)(b_j - m_2) + \mu^2 \sum_{i=1}^{m}\sum_{j=1}^{n} p_{ij}(b_j - m_2)^2 =$$

$$= \sigma_1^2\lambda^2 + 2\lambda\mu\rho\sigma_1\sigma_2 + \sigma_2^2\mu^2 \geqslant 0.$$

Forma pătratică considerată, fiind pozitiv definită, discriminantul ei este pozitiv.

Prin urmare

$$\begin{vmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{vmatrix} = \sigma_1^2\sigma_2^2(1 - \rho^2) \geqslant 0$$

rezultă

$$\rho^2 \leqslant 1$$

sau

$$-1 \leqslant \rho \leqslant 1.$$

2°. Dacă coeficientul de corelație $\rho = \pm 1$, între variabilele x și y ale distribuției există o dependență liniară.

Într-adevăr, discriminantul formei pătratice, de mai sus, pozitiv definit este nul. În acest caz este necesar și suficient ca parametrii λ și μ să verifice relația liniară

$$\lambda(a_i - m_1) + \mu(b_j - m_2) = 0, \quad (i, j = 1, 2, \ldots, m).$$

Valorile a_i și b_j luate de variabilele x și y ale distribuției verifică deci o relație liniară și prin urmare între cele două variabile există dependența liniară

$$\lambda(x - m_1) + \mu(y - m_2) = 0.$$

Pe dreapta definită de această ecuație se vor găsi toate punctele $P(a_i, b_j)$ ale distribuției; distribuția bidimensională se reduce la o distribuție unidimensională.

Parametrii λ și μ din ecuația precedentă se pot determina ușor. Discriminantul formei pătratice pozitiv definită fiind nul întrucît $\rho = \pm 1$, avem

$$\frac{\lambda}{\mu} = \frac{-\rho\sigma_1\sigma_2 \pm \sqrt{\sigma_1^2\sigma_2^2(\rho^2 - 1)}}{\sigma_1^2} = \mp\frac{\sigma_2}{\sigma_1}.$$

În mod analog,

$$\frac{\mu}{\lambda} = -\frac{\rho\,\sigma_1\,\sigma_2 \pm \sqrt{\sigma_1^2\,\sigma_2^2\,(\rho^2-1)}}{\sigma_2^2} = \mp\frac{\sigma_1}{\sigma_2}.$$

Obținem, astfel, următoarele drepte pe care se află punctele distribuției

$$y - m_2 = \frac{\sigma_2}{\sigma_1}(x - m_1) \qquad (\rho = 1)$$

$$y - m_2 = -\frac{\sigma_2}{\sigma_1}(x - m_1) \qquad (\rho = -1)$$

respectiv

$$x - m_1 = \frac{\sigma_1}{\sigma_2}(y - m_2) \qquad (\rho = 1)$$

$$x - m_1 = -\frac{\sigma_1}{\sigma_2}(y - m_2) \qquad (\rho = -1).$$

Acestea sînt dreptele de regresie (XI, 5.5) și (XI, 5.6) găsite mai înainte, în cazul particular cînd $\rho = \pm 1$.

Dacă ρ este foarte apropiat de ± 1 vom putea presupune distribuția examinată cu toate punctele apropiate de dreapta de regresie.

3°. Dacă variabilele x și y ale distribuției sînt independente între ele, atunci coeficientul de corelație este nul

$$\rho = 0.$$

Proprietatea rezultă imediat, deoarece în acest caz

$$M[(x - m_1)(y - m_2)] = M(x - m_1)\,M(y - m_2) = 0.$$

Dacă, însă, $\rho = 0$ nu putem trage concluzia că există o dependență slabă între variabilele x și y ale distribuției. Dimpotrivă, exemplul următor ne dovedește contrariul. Să considerăm următorul tablou de corelație.

y \ x	-1	0	1	
0	0	$\frac{1}{2}$	0	$\frac{1}{2}$
1	$\frac{1}{4}$	0	$\frac{1}{4}$	$\frac{1}{2}$
	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$	1

Fig. 27

Să observăm că punctele distribuției sînt așezate pe parabola $y = -x^2$; de asemenea că

$$m_1 = 0 \cdot \frac{1}{2} + 1 \cdot \frac{1}{2} = \frac{1}{2},$$

$$m_2 = (-1) \cdot \frac{1}{4} + 0 \cdot \frac{1}{2} + 1 \frac{1}{4} = 0,$$

$$M(x - m_1)(y - m_2) = M(x, y) - m_1 m_2 = (-1) \cdot \frac{1}{4} + \frac{1}{4} = 0$$

și, prin urmare,

$$\rho = 0.$$

În timp ce coeficientul de corelație este nul, între variabilele x și y dependența este directă, funcțională, $y = x^2$.

Cu atît mai mult, dacă $\rho = 0$ nu rezultă că variabilele x și y sînt independente între ele, cum se poate constata din exemplul următor. Fie o distribuție formată din opt puncte ca în figura 27.

În fiecare din punctele distribuției, masa p_{ij}, $(i, j = 1, 2, \ldots, 8)$, este egală evident, cu $\frac{1}{8}$.

Se constată ușor că

$$m_1 = m_2 = 0$$

$$M(x - m_1)(y - m_2) = 0$$

și deci

$$\rho = 0.$$

Calculînd masa din punctul $(2, 2)$, de pildă, găsim

$$p_{22} = \frac{1}{8}.$$

În același timp avem

$$p_{2\cdot} = \frac{1}{4}, \; p_{\cdot 2} = \frac{1}{4}$$

adică

$$p_{22} \neq p_{2\cdot} \, p_{\cdot 2}$$

Așadar, dacă coeficientul de corelație $\rho = 0$, nu reiese în mod necesar independența variabilelor x și y.

De aceea, cînd coeficientul de corelație este nul se spune că variabilele x și y ale distribuției sînt *necorelate*, ceea ce, în lumina exemplului precedent,

cînd între y și x este dependența funcțională $y = x^2$ nu poate reprezenta o calificare foarte precisă. Coeficientul de corelație ρ indică deci, numai într-o măsură ce nu poate fi considerată satisfăcătoare, legătura dintre variabilele x și y. Numai în cazul cînd el ia valorile limită ± 1 putem vorbi despre o situație categorică : există o dependență liniară între cele două variabile x și y.

Vom face acum o aplicație pentru a arăta concret modul de calcul al coeficientului de corelație liniară și al dreptei de regresie. În acest scop vom folosi datele tabelului 32, calculînd indicatorii necesari :

$$m_1 = M(x) = \frac{40 + 72 + 184 + 161}{58} = \frac{457}{58} = 7{,}88,$$

$$m_2 = M(y) = \frac{30 + 108 + 208 + 119}{58} = \frac{465}{58} = 8{,}02,$$

$$M_{20} = M(x^2) = \frac{4 \cdot 10^2 + 8 \cdot 9^2 + 23 \cdot 8^2 + 23 \cdot 7^2}{58} =$$

$$= \frac{400 + 648 + 1472 + 1127}{58} = \frac{3647}{58} = 62{,}88,$$

$$M_{02} = M(y^2) = \frac{3 \cdot 10^2 + 12 \cdot 9^2 + 26 \cdot 8^2 + 17 \cdot 7^2}{58} =$$

$$= \frac{300 + 972 + 1664 + 833}{58} = \frac{3769}{58} = 64{,}99,$$

$$M_{11} = M(x \cdot y) = \frac{2 \cdot 100 + 2 \cdot 90 + 1 \cdot 90 + 6 \cdot 81 + 1 \cdot 72 + 4 \cdot 72 + 15 \cdot 64 +}{58}$$

$$\frac{+ 4 \cdot 56 + 10 \cdot 56 + 13 \cdot 49}{58} =$$

$$= \frac{200 + 180 + 90 + 486 + 72 + 288 + 960 + 224 + 560 + 637}{58} =$$

$$= \frac{3697}{58} = 63{,}75,$$

$$(m_1)^2 = 7{,}88^2 = 62{,}09,$$

$$(m_2)^2 = 8{,}02^2 = 64{,}32,$$

$$m_1 m_2 = 63{,}20,$$

$$\sigma_x^2 = 62{,}88 - 62{,}09 = 0{,}79 \; ; \; \sigma_x = \sqrt{0{,}79} = 0{,}89,$$

$$\sigma_y^2 = 64{,}99 - 64{,}32 + 0{,}67 \; ; \; \sigma_y = \sqrt{0{,}67} = 0{,}82,$$

$$M(xy) - M(x) \cdot M(y) = 63{,}75 - 63{,}20 = 0{,}55.$$

Introducem acești indicatori în calculul coeficientului de corelație:

$$\rho = \frac{M(xy) - M(x) \cdot M(y)}{\sqrt{\sigma_x^2 \sigma_y^2}} = \frac{\overline{x \cdot y} - \bar{x} \cdot \bar{y}}{\sigma_x \sigma_y} = \frac{0{,}55}{0{,}73} = 0{,}75.$$

După cum se vede, valoarea coeficientului de corelație este destul de mare ceea ce este și explicabil. În general, elevii care au note bune la matematici au și note bune la fizică. Între varibilele x și y este o corelație directă destul de puternică.

Vom calcula acum dreptele de regresie, aplicînd formulele (XI, 5.5) și (XI, 5.6):

$$y - 8{,}02 = \frac{0{,}75 \cdot 0{,}82}{0{,}89}(x - 7{,}88),$$

$$y = 2{,}58 + 0{,}69\, x,$$

$$y - 8{,}02 = \frac{0{,}75 \cdot 0{,}89}{0{,}82}(x - 7{,}88),$$

$$y = 1{,}56 + 0{,}81\, x.$$

Se constată că cele două drepte trec prin centrul distribuției, adică prin punctul M ale cărei coordonate sînt $x = 7{,}88$, $y = 8{,}02$.

7. Corelația rangurilor

Să presupunem că avem o serie de n unități statistice $A_1, A_2, \ldots, A_n$, fiecare dintre ele avînd două caracteristici și anume respectiv

$$(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n).$$

De exemplu $A_1, A_2, \ldots, A_n$ pot reprezenta n persoane, $x_1, x_2, \ldots, x_n$, respectiv înălțimea lor măsurată în centimetri și $y_1, y_2, \ldots, y_n$ respectiv greutatea lor măsurată în kilograme. Se pune problema dacă există o corelație între aceste două caracteristici, înălțimea și greutatea.

Să admitem pentru simplificare că avem $n = 10$ persoane și că am făcut o clasificare a lor după cele două caracteristici. Această clasificare ne-a dat următoarele rezultate

A_1	A_2	A_3	A_4	A_5	A_6	A_7	A_8	A_9	A_{10}	
3	1	10	6	2	5	7	8	4	9	(α)
1	2	9	7	4	3	6	10	5	8	

În prima linie sînt așezate cele 10 persoane. În linia a doua ordinea pe care o ocupă fiecare persoană atunci cînd așezăm înălțimile lor în ordinea crescîndă iar în linia a treia aceeași ordine referitoare însă la greutatea corpului. De exemplu A_1 este al treilea în ordinea înălțimii dintre cele 10 persoane (primul fiind cel mai mărunt și ultimul cel mai înalt) și are greutatea cea mai mică dintre toți. A_2 ocupă locul întîi la înălțime, adică este cel mai mărunt și locul al doilea la greutatea corpului.

Este evident că această ordonare ne poate da un criteriu privind legătura dintre aceste două caracteristici.

De exemplu, dacă am fi avut o înșiruire de forma :

$$\left.\begin{array}{cccccccccc} A_1 & A_2 & A_3 & A_4 & A_5 & A_6 & A_7 & A_8 & A_9 & A_{10} \\ 3 & 1 & 10 & 6 & 2 & 5 & 7 & 8 & 4 & 9 \\ 3 & 1 & 10 & 6 & 2 & 5 & 7 & 8 & 4 & 9 \end{array}\right\} (\beta)$$

înseamnă că am fi avut cea mai strînsă dependență dintre cele două caracteristici. În acest caz, fiecare persoană $A_i, (i = 1, \ldots, 10)$, ocupă același loc în ordinea făcută după cele două caracteristici.

Dacă, dimpotrivă, am fi avut o ordonare de forma

$$\left.\begin{array}{cccccccccc} A_1 & A_2 & A_3 & A_4 & A_5 & A_6 & A_7 & A_8 & A_9 & A_{10} \\ 3 & 1 & 10 & 6 & 2 & 5 & 7 & 8 & 4 & 9 \\ 8 & 10 & 1 & 5 & 9 & 6 & 4 & 3 & 7 & 2 \end{array}\right\} (\gamma)$$

atunci am fi avut discordanța maximă între aceste două caracteristici. Se vede că în acest caz, persoana A_2 care este prima în ordonarea înălțimilor este ultima în ceea ce privește greutatea ș.a.m.d.

Rangurile obținute în cele două clasificări ne pot oferi deci o indicație asupra corelației dintre cele două caracteristici. Coeficientul de corelație astfel obținut poartă numele de *coeficient de corelație a rangurilor*.

Să ne referim la clasificarea (α). Diferențele dintre linia a doua și a treia sînt respectiv $2, -1, 1, -1, -2, 2, 1, -2, -1, 1$.

Dacă notăm prin d_i $(1 \leqslant i < 10)$ aceste diferențe, este evident că trebuie să avem

$$\sum_{i=1}^{10} d_i = 0.$$

Cu cît diferențele d_i sînt mai mari în valoare absolută, cu atît avem o mai mare discordanță între mărimile celor două caracteristici. Putem măsura atunci corelația dintre aceste două caracteristici folosind fie $\sum_i |d_i|$ — suma diferențelor în valoare absolută, fie $\sum_i d_i^2$ — suma pătratelor diferențelor. Din motive pe care le-am expus la măsurarea dispersiei, este preferabil să folosim $\sum_i d_i^2$.

Cea mai mare valoare a lui $\sum_i d_i^2$ obținem pentru distribuția de forma (γ). Vom avea în acest caz

$$\sum_i d_i^2 = \sum_i (x_i - y_i)^2 = \sum_i x_i^2 + \sum_i y_i^2 - 2\sum_i x_i y_i.$$

$$\sum_i x_i^2 = 1^2 + 2^2 + \ldots + n^2 = \frac{n(n+1)(2n+1)}{6}$$

$$\sum_i y_i^2 = 1^2 + 2^2 + \ldots + n^2 = \frac{n(n+1)(2n+1)}{6}$$

$$\begin{aligned}\sum_i x_i y_i &= 1 \cdot n + 2(n-1) + 3(n-2) + \ldots + n[n-(n-1)] = \\ &= 1 \cdot [(n+1)-1] + 2[(n+1)-2] + 3[(n+1)-3] + \\ &+ \ldots + n[(n+1)-n] = (1 + 2 + \ldots + n)(n+1) - \\ &- 1^2 - 2^2 \ldots n^2 = \frac{n(n+1)^2}{2} - \frac{n(n+1)(2n+1)}{6} = \\ &= \frac{n(n+1)(n+2)}{6}.\end{aligned}$$

Ținînd seamă de valorile obținute, rezultă

$$\sum_i d_i^2 = \frac{n(n+1)(2n+1)}{3} - \frac{n(n+1)(n+2)}{3} = \frac{n^3 - n}{3}.$$

În baza acestui rezultat, vom introduce coeficientul de corelație al rangurilor

$$\rho = 1 - \frac{6\sum_i d_i^2}{n(n^2-1)}. \qquad \text{(XI, 7.1)}$$

Acest indicator poartă numele de coeficientul lui Spearman.

$\rho = 1$ cînd avem maxima concordanță, $\rho = -1$ cînd avem maxima discordanță.

În cazul tabloului (α), avem

$$\sum_i d_i^2 = 22.$$

Prin urmare

$$\rho = 1 - \frac{6 \cdot 22}{10^3 - 10} = 0{,}867,$$

coeficientul ρ fiind destul de apropiat de 1, înseamnă că avem o corelație destul de strînsă între cele două caracteristici.

Procedînd în același mod ca mai sus se arată că

$$D^2(\rho) = \frac{1}{n+1}.$$

Un alt coeficient de corelație a rangurilor a fost dat de Kendall. Modul de construcție al coeficientului lui Kendall va rezulta mai ușor folosind tot un exemplu particular (α) :

$$\left.\begin{array}{cccccccccc} A_2 & A_5 & A_1 & A_9 & A_6 & A_4 & A_7 & A_8 & A_1 & A_3 \\ 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 \\ 2 & 4 & 1 & 5 & 3 & 7 & 6 & 10 & 8 & 9 \end{array}\right\} (\delta)$$

Tabloul (δ) conține aceleași elemente ca și tabloul (α), cu singura diferență că în linia a doua, rangurile au fost scrise în ordine naturală, de la 1 la 10. Fiecare persoană A_k, ($k = 1, 2, \ldots, 10$), are însă aceleași caracteristici în ambele tablouri.

Să considerăm prima persoană din tabloul (δ), adică persoana A_2. Această persoană ocupă locul 1 în ceea ce privește înălțimea și locul 2 în ceea ce privește greutatea. Prin urmare, există 9 persoane mai înalte decît A_2 și numai 8 mai grele. Dacă ar fi fost o perfectă concordanță între cele două caracteristici, ar fi trebuit să fie și 9 persoane mai grele. Dar sînt numai 8. Deci nu există o perfectă concordanță. Totuși faptul că sînt 8 persoane mai grele arată oarecum că există o legătură între cele două caracteristici în sensul că a doua crește în intensitate, cînd și prima crește. N-am fi putut afirma același lucru, cînd în locul numărului 8 am fi avut un număr cu mult mai mic.

Deci numărul numerelor mai mari decît 2 în linia a treia reprezintă pentru noi un indicator al concordanței pozitive.

A doua persoană A_5 este a doua în ceea ce privește înălțimea și a patra în ceea ce privește greutatea. Sînt 8 persoane mai înalte decît A_5 și numai 6 mai grele. 6 este indicatorul concordanței pozitive pentru A_5.

Procedînd astfel pentru toate persoanele A_k, găsim următorii indicatori :

$$8, 6, 7, 5, 5, 3, 3, 0, 1.$$

Suma acestor indicatori, pe care o notăm prin P, ne dă

$$P = 38.$$

Dacă am fi avut maximul concordanței pozitive, am fi avut indicatorii

$$9, 8, 7, 6, 5, 4, 3, 2, 1,$$

pentru care $P = 45$.

Pentru n termeni, maximul concordanței pozitive ar fi dat pentru P valoarea

$$P = (n-1) + (n-2) + \ldots + 2 + 1 = \frac{n(n-1)}{2}$$

Dar mai putem judeca și într-alt mod. Prima persoană A_2 nu are nici o altă persoană mai măruntă decît ea, în schimb există una mai ușoară și anume A_1. Dacă ar fi existat perfectă concordanță pozitivă între înălțime și greutate, nu ar fi trebuit să existe nici o persoană mai ușoară.

Faptul că există o persoană mai ușoară este un indicator al discordanței, căci dacă numărul persoanelor mai ușoare ar fi fost și mai mare am fi avut o discordanță și mai mare.

Procedăm la fel pentru fiecare persoană A_k, numărînd cîte numere sînt mai mici la dreapta lui A_k, în linia a treia, decît caracteristica ce corespunde lui A_k.

Găsim următoarele numere :

$$1,\ 2,\ 0,\ 1,\ 0,\ 1,\ 0,\ 2,\ 0$$

iar suma lor

$$Q = 7.$$

Discordanța maximă ar fi fost în cazul cînd am fi avut indicatorii :

$$9,\ 8,\ 7,\ 6,\ 5,\ 4,\ 3,\ 2,\ 1$$

și în acest caz

$$Q = 45.$$

Dacă am fi avut o serie de n termeni, discordanța maximă ar fi corespuns șirului

$$(n-1),\ (n-2),\ \ldots,\ 2,\ 1$$

și

$$Q = \frac{n(n-1)}{2}.$$

Obținem deci un indicator al concordanței, considerînd raportul

$$\rho = \frac{2(P-Q)}{n(n-1)}.$$

Cînd acest raport este egal cu 1, avem concordanța maximă, cînd este egal cu -1, discordanța maximă, iar cînd este egal cu zero, independența între cele două caracteristici.

În cazul nostru :

$$\rho = \frac{2(38-7)}{10 \cdot 9} = 0{,}69,$$

ceea ce indică o concordanță pozitivă.

Dispersia lui ρ este dată de formula

$$D^2(\rho) = \frac{2(2n+5)}{9n(n-1)}.$$

Pentru a putea trage concluzii valabile asupra legăturii dintre cele două caracteristici este necesar să se țină seama de mărimea dispersiei și să se construiască cu ajutorul ei un interval, care cu probabilitate dată, să cuprindă adevărata valoare a coeficientului de corelație a rangului. Dispersia este destul de mare pentru valorile mici ale lui n. De aceea aplicarea acestui coeficient pentru serii mici de valori nu este indicată.

Scanned with OKEN Scanner

8. Distribuții cu trei dimensiuni

Pentru colectivitățile statistice cu trei dimensiuni se poate întocmi un tablou de corelație, întocmai ca și pentru colectivitățile cu două caracteristici. Vom pleca de la un exemplu simplu de tablou de corelație cu trei dimensiuni. Să considerăm o colectivitate formată din 300 de studenți din anul II ai unei facultăți de matematică, pe care îi distribuim după notele luate în anul I la examenele de algebră, calcul diferențial și integral și geometrie analitică. Avem trei caracteristici :

$x =$ nota la examenul de algebră ;
$y =$ „ „ „ „ calcul diferențial și integral ;
$z =$ „ „ „ „ geometrie analitică.

Fiecare din aceste caracteristici poate lua valorile 5, 6, 7, 8, 9 și 10.

Notăm prin f_{xyz} frecvența corespunzătoare valorilor x, y, z care reprezintă în cazul nostru numărul studenților care au luat nota x la algebră, y la calcul diferențial și integral, z la geometrie analitică.

Alegînd un sistem de trei axe de coordonate Ox, Oy, Oz în spațiul cu trei dimensiuni, vom reprezenta valorile x, y, z respectiv pe cele trei axe Ox, Oy, Oz. Vom împărți spațiul prin planuri paralele în celule formate din cuburi, întocmai după cum în cazul corelației în plan, am împărțit planul în celule formate din pătrate, ducînd paralele la axele de coordonate (vezi fig. 28).

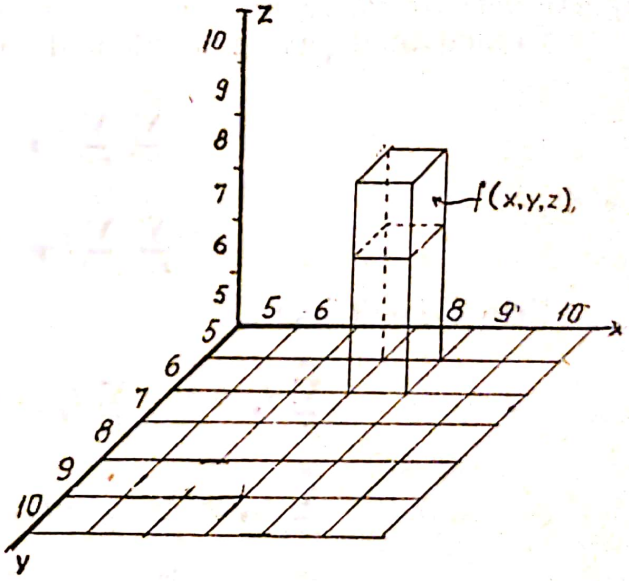


Fig. 28

Pe axa Ox am așezat notele 5, 6, 7, 8, 9, 10 ; la fel și pe celelalte două axe Oy, Oz.

Pentru simplificare, pe figura 28, se găsește desenat cubul corespunzător valorilor $x = 6$, $y = 8$, $z = 7$. În interiorul acestui cub, scriem frecvența $f_{6,8,7}$, care ne arată numărul studenților care au luat nota 6 la algebră, nota 8 la calcul diferențial și integral, nota 7 la geometria analitică.

Este evident că

$$\sum_{x=5}^{10} \sum_{y=5}^{10} \sum_{z=5}^{10} f_{xyz} = N,$$

este frecvența totală, iar

$$\sum_{x=5}^{10} f_{xyz} = f_{yz}$$

reprezintă frecvența corespunzătoare valorilor y și z, adică în cazul nostru numărul studenților care au luat nota y la calcul diferențial și integral și nota z la geometrie analitică.

f_{yz} reprezintă o distribuție bidimensională, cu caracteristicile y și z. Ea poartă numele de distribuție marginală.

În mod analog, din distribuția tridimensională extragem distribuțiile bidimensionale marginale:

$$\sum_{y=5}^{10} f_{xyz} = f_{xz},$$

$$\sum_{z=5}^{10} f_{xyz} = f_{xy}.$$

Să considerăm suma

$$\sum_{x=5}^{10} \sum_{y=5}^{10} f_{xyz} = f_z,$$

f_z reprezintă numărul de studenți care au luat nota z la geometrie analitică, f_z este deci frecvența unei distribuții marginale unidimensionale.

În mod analog avem alte două distribuții marginale unidimensionale:

$$\sum_{y=5}^{10} \sum_{z=5}^{10} f_{xyz} = f_x,$$

$$\sum_{z=5}^{10} \sum_{x=5}^{10} f_{xyz} = f_y.$$

Este evident că:

$$\sum_{x=5}^{10} f_x = N, \quad \sum_{y=5}^{10} f_y = N, \quad \sum_{z=5}^{10} f_z = N,$$

$$\sum_{y=5}^{10} f_{xy} = \sum_{z=5}^{10} f_{xz} = f_x,$$

$$\sum_{z=5}^{10} f_{yz} = \sum_{x=5}^{10} f_{xy} = f_y,$$

$$\sum_{x=5}^{10} f_{xz} = \sum_{y=5}^{10} f_{yz} = f_z.$$

Frecvențele f_{xyz} sînt trecute în tabelul 40. Pentru simplificare am folosit o nouă grupare. Notele 5 și 6 le-am înlocuit cu 5,5, notele 7 și 8 prin 7,5, notele 9 și 10 prin 9,5.

Problema corelației și a regresiunii se rezolvă pentru acest tabel cu trei serii de date în mod analog ca și pentru cazul cînd avem numai două serii de date.

Distribuții tridimensionale. **Definiția matematică a distribuțiilor teoretice și empirice constituie o extindere imediată a noțiunilor întîlnite pentru distribuțiile cu o singură sau două dimensiuni.**

Să considerăm vectorul $V(\xi, \eta, \zeta)$, prin ξ, η, ζ notînd componentele lui, care sînt variabile aleatoare, într-un spațiu euclidean R_3, cu trei dimensiuni.

Fie $V_1(a_1, b_1, c_1)$, $V_2(a_2, b_2, c_2)$ doi vectori dați prin componentele lor.

Tabelul 40

Nr. crt.	x	y	x	f_{xyz}
1	5,5	5,5	5,5	70
2	5,5	5,5	7,5	14
3	5,5	5,5	9,5	1
4	5,5	7,5	5,5	14
5	5,5	7,5	7,5	7
6	9,5	9,5	7,5	2
7	5,5	9,5	9,5	1
8	7,5	5,5	5,5	21
9	7,5	5,5	7,5	8
10	7,5	5,5	9,5	4
11	7,5	7,5	5,5	16
12	7,5	7,5	7,5	23
13	7,5	7,5	9,5	4
14	7,5	9,5	5,5	6
15	7,5	9,5	7,5	8
16	7,5	9,5	9,5	6
17	9,5	5,5	5,5	4
18	9,5	5,5	7,5	8
19	9,5	5,5	9,5	1
20	9,5	7,5	5,5	2
21	9,5	7,5	7,5	10
22	9,5	7,5	9,5	10
23	9,5	9,5	5,5	3
24	9,5	9,5	7,5	14
25	9,5	9,5	9,5	43

Să considerăm probabilitatea inegalităților

$$a_1 < \xi \leqslant a_2,\ b_1 < \eta \leqslant b_2,\ c_1 < \zeta \leqslant c_2,$$

pe care o notăm prin $P(V_1 < V \leqslant V_2)$.

Considerăm funcția $F(x,y,z)$, de variabilele reale x, y, z, care se bucură de următoarele proprietăți:

$$0 \leqslant F(x, y, z) \leqslant 1,$$

$$F(-\infty, y, z) = F(x_1 - \infty, z) = F(x_1, y_1) - \infty) = 0,$$

$$F(+\infty, +\infty, +\infty) = 1,$$

$$F(x + h, y + k, z + l) - F(x, y + k, z + l) - F(x + h, y, z + l) - F(x + h, y + k, z) + F(x, y, z + l) + F(x, y + k, z) + F(x + h, y, z) - F(x, y, z) \geqslant 0.$$

$F(x, y, z)$ este nedescrescătoare în raport cu fiecare variabilă x, y, z luată în parte și continuă la dreapta.

O funcție $F(x, y, z)$ care îndeplinește aceste condiții poartă numele de funcție de repartiție și definește în mod unic o lege de probabilitate cu trei dimensiuni.

Scanned with OKEN Scanner

Probabilitatea $P(V_1 < V \leqslant V_2)$ se exprimă cu ajutorul lui $F(x, y, z)$ în modul următor:

$$P(V_1 < V \leqslant V_2) = F(a_2, b_2, c_2) - F(a_1, b_2, c_2) - F(a_2, b_1, c_2) - F(a_2, b_2, c_1) + F(a_1, b_1, c_2) + F(a_1, b_2, c_1) + F(a_2, b_1, c_1) - F(a_1, b_1, c_1).$$

Dacă există derivata parțială

$$\frac{\partial^3 F(x, y, z)}{\partial x \partial y \partial z} = f(x, y, z)$$

în orice punct x, y, z, funcția $f(x, y, z)$ fiind continuă în raport cu x, y, z, legea de probabilitate este continuă, iar $f(x, y, z)$ poartă numele de densitate de probabilitate.

Legea de probabilitate este discretă, atunci cînd este compusă dintr-un număr finit de puncte, care pot fi luate cu probabilități date. De asemenea sîntem în cazul discret, cînd există o infinitate numărabilă de puncte, cu probabilitățile respective, cu condiția ca în orice domeniu mărginit să se găsească un număr finit de puncte.

Distribuțiile empirice tridimensionale se definesc în același mod, cu diferența că în loc de probabilitatea $P(V_1 < V \leqslant V_2)$ introducem frecvența corespunzătoare $M(V_1 < V \leqslant V_2)$, iar funcția $F(x, y, z)$, îndeplinind aceleași condiții ca și funcția de repartiție, poartă numele de *funcție de frecvență cumulată sau ogivă*.

Momentele de primul ordin sînt date de egalitățile:

$$m_1 = \iiint_{R_3} x\,dF(x, y, z),\quad m_2 = \iiint_{R_3} y\,dF(x, y, z),\quad m_3 = \iiint_{R_3} z\,dF(x, y, z)$$

integralele triple fiind extinse la tot spațiul R_3.

Momentele de ordinul al doilea sînt:

$$M_{11} = \iiint_{R_3} x^2\,dF(x, y, z),\quad M_{22} = \iiint_{R_3} y^2\,dF(x, y, z),\quad M_{33} = \iiint_{R_3} z^2\,dF(x, y, z),$$

$$M_{12} = M_{21} = \iiint_{R_3} xy\,dF(x, y, z),\quad M_{23} = M_{32} = \iiint_{R_3} yz\,dF(x, y, z),$$

$$M_{31} = M_{13} = \iiint_{R_3} xz\,dF(x, y, z).$$

Rezultă imediat momentele centrate de ordinul al doilea, date de formulele:

$$m_{11} = M_{11} - m_1^2,\quad m_{22} = M_{22} - m_2^2,\quad m_{33} = M_{33} - m_3^2,$$

$$m_{12} = m_{21} = M_{12} - m_1 m_2,\quad m_{32} = m_{23} = M_{23} - m_2 m_3,$$

$$m_{31} = m_{13} = M_{13} - m_1 m_3.$$

Ele rezultă din formulele care exprimă momentele de ordinul al doilea raportate respectiv la valorile medii m_{11}, m_{21}, m_{31}.

Avem astfel:

$$m_{12} = \iiint_{R_3} (x - m_1)(y - m_2)\,dF(x, y, z) = \iiint_{R_3} xy\,dF(x, y, z) - m_1 \iiint_{R_3} y\,dF(x, y, z) - m_2 \iiint_{R_3} x\,dF(x, y, z) + m_1 m_2 \iiint_{R_3} dF(x, y, z)$$

unde $\iiint_{R_3} dF(x, y, z) = M_{12} - m_1 m_2.$

Celelalte formule se stabilesc în mod analog.

Vom mai introduce notațiile:

$$m_{11} = \sigma_1^2,\ m_{22} = \sigma_2^2,\ m_{33} = \sigma_3^2,$$

$$m_{12} = \rho_{12}\sigma_1\sigma_2,\ m_{23} = \rho_{23}\sigma_2\sigma_3,\ m_{31} = \rho_{31}\sigma_3\sigma_1.$$

Se vede că σ_1^2, σ_2^2, σ_3^2, reprezintă respectiv dispersiile componentelor ξ, η, ζ, pe cînd ρ_{12}, ρ_{23}, ρ_{31}, coeficienții de corelație ale variabilelor ξ, η ζ luate două cîte două.

Este util să considerăm matricele:

$$M = \begin{Vmatrix} m_{11} & m_{12} & m_{13} \\ m_{21} & m_{22} & m_{23} \\ m_{31} & m_{32} & m_{33} \end{Vmatrix}$$

$$N = \begin{Vmatrix} 1 & \rho_{12} & \rho_{13} \\ \rho_{21} & 1 & \rho_{23} \\ \rho_{31} & \rho_{32} & 1 \end{Vmatrix}$$

9. Problema regresiei

Să considerăm forma liniară:

$$\alpha + \beta y + \gamma z.$$

Ne propunem să determinăm parametrii α, β, σ, astfel ca expresia:

$$\iiint_{R_3} x - \alpha - \beta y - \gamma z)^2\, dF(x, y, z), \qquad \text{(XI, 9.1)}$$

să fie minimă.

După cum se vede, problema revine la a face minimă valoarea medie a expresie $(x - \alpha - \beta y - \gamma z)^2$, pentru punctele (x, y, z) ale distribuției statistice considerate.

Pentru a obține minimul integralei (XI, 9.1) este necesar s-o derivăm respectiv în raport cu α, β, γ.

Anulînd derivata ei parțială de ordinul întîi în raport cu α, avem cu notațiile adoptate:

$$\alpha + m_2\beta + m_3\gamma = m_1.$$

Deducem

$$\alpha = m_1 - m_2\beta - m_3\gamma.$$

Introducînd această valoare în (XI, 9.1) trebuie să facem minimă integrala:

$$\iiint_{R_3} [(x - m_1 - \beta(y - m_2) - \gamma(z - m_3)]^2\, dF(x, y, z).$$

Derivînd acum în raport cu β și γ, obținem

$$m_{22}\beta + m_{23}\gamma = m_{12},$$

$$m_{32}\beta + m_{33}\gamma = m_{13}.$$

Rezolvarea acestui sistem ne dă:

$$\beta = \frac{\begin{vmatrix} m_{12} & m_{23} \\ m_{13} & m_{33} \end{vmatrix}}{\begin{vmatrix} m_{22} & m_{23} \\ m_{32} & m_{33} \end{vmatrix}}, \qquad \gamma = \frac{\begin{vmatrix} m_{22} & m_{12} \\ m_{32} & m_{13} \end{vmatrix}}{\begin{vmatrix} m_{22} & m_{23} \\ m_{32} & m_{33} \end{vmatrix}}.$$

Să considerăm determinantul matricei M:

$$M = \begin{vmatrix} m_{11} & m_{12} & m_{13} \\ m_{21} & m_{22} & m_{23} \\ m_3 & m_{32} & m_{33} \end{vmatrix}$$

și să notăm prin $\overline{M}_{ij}$ minorul elementului m_{ij} din determinantul $|M|$.

Cu această notație, valorile mai sus găsite pentru β și γ pot fi scrise:

$$\beta = -\frac{\overline{M}_{12}}{\overline{M}_{11}}, \quad \gamma = -\frac{\overline{M}_{13}}{\overline{M}_{11}}. \tag{XI, 9.2}$$

Coeficienții β și γ poartă numele de coeficienți de regresie.

Planul

$$\overline{M}_{11}(x - m_1) + \overline{M}_{12}(y - m_2) + \overline{M}_{13}(z - m_3) = 0, \tag{XI, 9.3}$$

poartă numele de plan de regresie pentru variabila x. El se bucură de o proprietate importantă, pe care o vom pune în evidență.

Fie $P(x, y, z)$ un punct al distribuției statistice și fie $P'(x', y, z)$ punctul din planul (XI, 9.3), care are aceleași coordonate y și z ca și punctul considerat al distribuției. Deducem din (XI, 9.3):

$$x' = m_1 - \frac{\overline{M}_{12}}{\overline{M}_{11}}(y - m_2) - \frac{\overline{M}_{13}}{\overline{M}_{11}}(z - m_3). \tag{XI, 9.4}$$

Rezultă

$$x - x' = x - m_1 + \frac{\overline{M}_{12}}{\overline{M}_{11}}(y - m_2) + \frac{\overline{M}_{13}}{\overline{M}_{11}}(z - m_3)$$

sau

$$M(x - x')^2 = M[x - m_1 - \beta(y - m_2) - \gamma(z - m_3)]^2, \tag{XI, 9.5}$$

unde β și γ au valorile (XI, 9.2). Pentru aceste valori ale lui β și γ știm însă că (XI, 9.5) ia valoarea minimă. Prin urmare, planul de regresie al lui x se bucură de proprietatea că valoarea medie a pătratelor distanțelor, de la orice punct $P(x, y, z)$ al distribuției date la acest plan, luate paralel cu axa Ox, este minimă.

Se mai constată din (XI, 9.3) că planul de regresie trece prin centrul distribuției:

Planurile:

$$\begin{aligned} \overline{M}_{21}(x - m_1) + \overline{M}_{22}(y - m_2) + \overline{M}_{23}(z - m_3) = 0, \\ \overline{M}_{31}(x - m_1) + \overline{M}_{32}(y - m_2) + \overline{M}_{33}(z - m_3) = 0, \end{aligned} \tag{XI, 9.6}$$

reprezintă planurile de regresie respectiv ale lui y și z.

10. Corelație parțială și multiplă

Din distribuția tridimensională extragem o distribuție marginală bidimensională, referindu-ne la valorile x, y, z luate două cîte două.

Dacă ne referim la x și y, putem scrie

$$m_{11} = M(x - m_1)^2, \quad m_{12} = M(x - m_1)(y - m_2), \quad m_{22} = M(y - m_2)^2.$$

Cu ajutorul acestor date, obținem coeficientul de corelație parțială între x și y, pe care îl notăm prin ρ_{12}:

$$\rho_{12} = \frac{m_{12}}{\sqrt{m_{11} m_{22}}}. \qquad \text{(XI, 10.1)}$$

În mod analog avem și ceilalți coeficienți de corelație ρ_{23} și ρ_{31}. ρ_{12} arată corelația între x și y. S-ar putea însă să ne intereseze corelația dintre x și ansamblul valorilor y, z. Atunci folosim coeficientul multiplu de corelație, care măsoară corelația între x și x', dat de formula (XI, 9.4).

Coeficientul multiplu de corelație al variabilelor x și x' îl notăm prin $\rho_{1.23}$. Prin definiție avem

$$\rho_{1.23} = \frac{M[(x - m_1)(x' - m_1)]}{\sqrt{M(x - m_1)^2 M(x' - m_1)^2}}.$$

Avem

$$M(x - m_1)^2 = m_{11}$$

$$M[(x - m_1)(x' - m_1)] = M\left\{(x - m_1)\left(-\frac{\overline{M}_{12}}{\overline{M}_{11}}(y - m_2) - \frac{\overline{M}_{13}}{\overline{M}_{11}}(z - m_3)\right)\right\} =$$

$$= -\frac{\overline{M}_{12}}{\overline{M}_{11}} M(x - m_1)(y - m_2) - \frac{\overline{M}_{13}}{\overline{M}_{11}} M(x - m_1)(z - m_3) = -\frac{\overline{M}_{12}}{\overline{M}_{11}} m_{12} -$$

$$- \frac{\overline{M}_{13}}{M_{11}} m_{13} = (m_{11}\overline{M}_{11} - |M|)\frac{1}{\overline{M}_{11}} = m_{11} - \frac{|M|}{\overline{M}_{11}},$$

deoarece

$$m_{11}\overline{M}_{11} + m_{12}\overline{M}_{12} + m_{13}\overline{M}_{13} = |M|$$

$$M(x' - m_1)^2 = M\left[-\frac{\overline{M}_{12}}{\overline{M}_{11}}(y - m_2) - \frac{\overline{M}_{13}}{\overline{M}_{11}}(z - m_3)\right]^2 =$$

$$= M\left[\frac{\overline{M}_{12}^2}{\overline{M}_{11}^2}(y - m_2)^2 + \frac{2\,\overline{M}_{12}\overline{M}_{13}}{\overline{M}_{11}^2}(y - m_2)(z - m_3) + \frac{\overline{M}_{13}^2}{\overline{M}_{11}^2}(z - m_3)^2\right] =$$

$$= \frac{m_{22}\overline{M}_{12}^2 + 2m_{23}\overline{M}_{12}\overline{M}_{13} + m_{33}\overline{M}_{13}^2}{\overline{M}_{11}^2}.$$

Pe de altă parte,

$$m_{22}\overline{M}_{12}^2 + 2m_{23}\overline{M}_{12}\overline{M}_{13} + m_{33}\overline{M}_{13}^2 = \overline{M}_{12}(m_{22}\overline{M}_{12} + m_{23}\overline{M}_{13}) +$$

$$+ \overline{M}_{13}(m_{23}\overline{M}_{12} + m_{33}\overline{M}_{13}) = \overline{M}_{12}(-m_{21}\overline{M}_{11}) + \overline{M}_{13}(-m_{13}\overline{M}_{13}) =$$

$$= -\overline{M}_{11}(m_{12}\overline{M}_{12} + m_{13}\overline{M}_{13}) = -\overline{M}_{11}[|\overline{M}| - m_{11}\overline{M}_{11}],$$

Scanned with OKEN Scanner

deoarece:

$$m_{21}\overline{M}_{11} + m_{22}\overline{M}_{12} + m_{23}\overline{M}_{13} = 0,$$
$$m_{13}\overline{M}_{13} + m_{23}\overline{M}_{12} + m_{33}\overline{M}_{13} = 0.$$

Prin urmare,

$$M(x' - m_1)^2 = m_{11} - \frac{|M|}{\overline{M}_{11}}$$

$$\rho_{1.23} = \frac{m_{11} - \dfrac{|M|}{\overline{M}_{11}}}{\sqrt{m_{11}\left(m_{11} - \dfrac{|M|}{\overline{M}_{11}}\right)}}, \qquad \text{(XI, 10.2)}$$

$$\rho_{1.23} = \sqrt{1 - \frac{|M|}{m_{11}\overline{M}_{11}}}.$$

În mod analog:

$$\rho_{2.31} = \sqrt{1 - \frac{|M|}{m_{22}\overline{M}_{22}}}, \qquad \text{(XI, 10.3)}$$

$$\rho_{3.12} = \sqrt{1 - \frac{|M|}{m_{33}\overline{M}_{33}}}.$$

Coeficientul multiplu de corelație este cuprins între 0 și 1. Dacă este egal cu 1, toate frecvențele se găsesc într-un plan, în loc să fie răspîndite în tot spațiul. Dacă este egal cu zero, de exemplu $\rho_{1.23} = 0$, înseamnă că x nu este corelat nici cu y, nici cu z.

EXEMPLUL 35. Vom aplica aceste considerații teoretice la exemplul dat în tabloul notelor medii de mai sus.

Făcînd calculele avem:

$$m_1 = 7{,}41, \quad m_2 = 7{,}15, \quad m_3 = 7{,}03.$$

Acestea sînt mediile generale obținute de studenți, respectiv la cele trei examene:

$$M_{11} = 66{,}42, \quad M_{22} = 58{,}54, \quad M_{33} = 63{,}23,$$
$$M_{12} = 59{,}82, \quad M_{23} = 58{,}52, \quad M_{31} = 62{,}05,$$

Relațiile de mai sus ne dau momentele absolute de ordinul al doilea. Din ele deducem valorile momentelor centrate de ordinul al doilea:

$$m_{11} = 11{,}51, \quad m_{22} = 7{,}42, \quad m_{33} = 13{,}81,$$
$$m_{12} = 6{,}84, \quad m_{23} = 8{,}26, \quad m_{31} = 9{,}96.$$

Scanned with OKEN Scanner

Vom scrie acum matricea M:

$$M = \begin{vmatrix} 11,51 & 6,84 & 9,96 \\ 6,84 & 7,42 & 8,26 \\ 9,96 & 8,26 & 13,81 \end{vmatrix}$$

al cărei determinant are valoarea:

$$|M| = \begin{vmatrix} 11,51 & 6,84 & 9,96 \\ 6,84 & 7,42 & 8,26 \\ 9,96 & 8,26 & 13,81 \end{vmatrix} = 137,4.$$

Minorii acestui determinant sînt respectiv:

$$\overline{M}_{11} = 34,24, \quad \overline{M}_{22} = 59,75, \quad \overline{M}_{33} = 38,61,$$

$$\overline{M}_{12} = \overline{M}_{21} = -12,19, \quad \overline{M}_{23} = \overline{M}_{32} = -26,94, \quad \overline{M}_{31} = \overline{M}_{13} = -17,40.$$

Aplicînd formulele (XI, 9.3) și (XI, 9.5), obținem ecuațiile planurilor de regresie pentru fiecare din cele trei valori x, y, z:

$$34,24\ x - 12,19\ y - 17,40\ z - 44,59 = 0,$$

$$-12,19\ x + 59,75\ y - 26,94\ z - 147,51 = 0,$$

$$-17,40\ x - 26,94\ y + 38,61\ z + 50,13 = 0,$$

pe care e mai convenabil să le scriem sub forma:

$$x = 0,36\ y + 0,51\ z + 1,30,$$

$$y = 0,20\ x + 0,45\ z + 2,47,$$

$$z = 0,45\ x + 0,70\ z - 1,30.$$

Se verifică lesne funcția planurilor de regresie. Să considerăm primul plan, care reprezintă regresia lui x. Să ne alegem valorile $y = 7,5$, $z = 7,5$. Înlocuind aceste valori în primul plan găsim $x = 7,83$. Din tabelul mediilor (nr. 40) rezultă că pentru $y = 7,5$, $z = 7,5$, găsim următoarele valori ale lui x, cu frecvențele corespunzătoare:

x	z
5,5	7
7,5	23
9,5	10
	40

Media valorilor lui x este 7,65, care se apropie foarte mult de valoarea $x = 7,83$, din planul de regresie. Se vede astfel modul în care planul de regresie se apropie cel mai mult, dintre toate planurile din spațiu, de distribuția valorilor x, în sensul pe care l-am arătat.

Coeficientul obișnuit de corelație între x și y este :

$$\rho_{12} = \frac{m_{12}}{\sqrt{m_{11}m_{22}}} = \frac{6,84}{\sqrt{11,51 \times 7,42}} = 0,74.$$

Se vede că avem o corelație pozitivă, corespunzătoare unui coeficient de corelație destul de mare. Ea denotă legătura dintre notele luate la examenele de geometrie analitică și algebră de același student.

În mod analog, obținem

$$\rho_{13} = \frac{m_{13}}{\sqrt{m_{11}m_{33}}} = \frac{9,96}{\sqrt{11,51 \times 13,81}} = 0,79,$$

$$\rho_{23} = \frac{m_{23}}{\sqrt{m_{22}m_{33}}} = \frac{8,26}{\sqrt{7,42 \times 13,81}} = 0,82.$$

Deci corelații pozitive, cu coeficienți de corelație ridicați, găsim și între x și z și între y și z.

Să trecem acum la coeficientul multiplu de corelație.

Coeficientul multiplu de corelație a lui x față de y și z este dat de formula (XI, 10.2) :

$$\rho_{1.23} = \sqrt{1 - \frac{|M|}{m_{11}\overline{M}_{11}}} = \sqrt{1 - \frac{137,4}{11,51 \times 66,42}} = 0,91.$$

Prin urmare, coeficientul de corelație al lui x numai față de y sau numai față de z, ia respectiv valorile

$$\rho_{12} = 0,74, \quad \rho_{13} = 0,79,$$

pe cînd coeficientul multiplu de corelație, față de ansamblul valorilor y și z, ia valoarea

$$\rho_{1.23} = 0,91,$$

care e cu mult mai mare.

În mod analog, avem :

$$\rho_{2.13} = \sqrt{1 - \frac{|M|}{m_{22}\overline{M}_{22}}} = \sqrt{1 - \frac{137,4}{7,42 \cdot 58,54}} = 0,82,$$

$$\rho_{3.12} = \sqrt{1 - \frac{|M|}{m_{33}M_{33}}} = \sqrt{1 - \frac{137,4}{13,81 \cdot 63,23}} = 0,92.$$

Se constată același lucru și anume că toți coeficienții multipli de corelație sînt mai mari decît coeficienții parțiali de corelație.

Coeficienții parțiali ρ_{12}, ρ_{23}, ρ_{31} arată corelația dintre notele luate la examene, luînd examenele două cîte două. Această corelație este însă mult mai puternică dacă se are în vedere ansamblul tuturor notelor, după cum se vede utilizînd coeficienții multipli de corelație.

Scanned with OKEN Scanner

11. Distribuții cu n dimensiuni

În cazul cînd avem n variabile, formulele stabilite mai sus pentru regresie și corelație se păstrează.

Fie $x_1, x_2, \ldots, x_n$ coordonatele unui punct P din spațiul R_n, cu n dimensiuni. Distribuția statistică în acest spațiu se definește în același mod ca în spațiul R_3.

Notăm prin M_{ij}, $(i, j = 1,2, \ldots, n)$, momentele absolute și prin m_{ij}, $(i, j = 1,2, \ldots, n)$, momentele centrate. Ecuația planului de regresie al variabilei x_i este:

$$\sum_{j=1}^{n} \overline{M}_{ij}(x_j - m_j) = 0, \qquad (i=1,2, \ldots, n).$$

Coeficientul multiplu de corelație al variabilei x_i față de celelalte variabile $x_1 \cdot x_2 \cdot x_{i-1}\, x_{i+1} \cdot x_n$, este dat de formula

$$\rho_{i,\,1,2,\ldots,\,i-1,\,i+1,\ldots,\,n} = \sqrt{1 - \frac{|M|}{m_{ii}\overline{M}_{ii}}},$$

unde $|M|$ și $\overline{M}_{ij}$ au semnificații analoage cu notațiile corespunzătoare din cazul a trei variabile.

PARTEA A TREIA

ANALIZA SERIILOR ÎN TIMP (DINAMICE)

Scanned with OKEN Scanner

Scanned with OKEN Scanner

Capitolul XII

Probleme generale ale elaborării şi analizei seriilor dinamice

1. Oscilaţiile în timp ale variabilelor şi construirea seriilor dinamice

Tratînd distribuţiile de frecvenţă, am văzut că teoria lor avea la bază o anumită *stabilitate*, un element de permanenţă în structura lor. Pe această stabilitate reală, obiectivă, am văzut că s-a construit şi teoria modernă a calculului probabilităţilor, a legilor de probabilitate care caracterizează diferite distribuţii teoretice. Tot atunci am precizat sensul legăturii dialectice dintre distribuţiile statistice ale colectivităţilor studiate şi distribuţiile teoretice corespunzătoare.

Este evident că vorbind despre colectivităţi şi unităţile individuale ce le compun, noi le-am considerat pînă acum, într-un moment dat al timpului. Mişcarea, schimbarea neîntreruptă a fenomenelor şi proceselor care au loc în natură şi în viaţa socială nu ne-au preocupat pînă acum direct, din punctul de vedere al cercetării ştiinţifice.

Acest fapt este desigur consecinţa naturii însăşi a problemelor legate de studiul statistic al mişcării, al dezvoltării dialectice a fenomenelor şi proceselor vieţii social-economice, care au un specific calitativ deosebit.

Este clar însă că în metodologia statisticii trebuie să existe anumite metode sau procedee de prelucrare tehnică a datelor reale, culese ştiinţific, oglindind dezvoltarea dialectică a fenomenelor şi proceselor social-economice pe care ea le studiază. În statistică s-au elaborat, într-adevăr, asemenea metode şi procedee tehnice fundamentate teoretic, pe cunoaşterea şi aplicarea concretă a celor mai generale legi ale dezvoltării întregii realităţi, legi pe care le dă dialectica materialistă, marxistă.

Este vorba de metode şi procedee de măsurare şi caracterizare statistică a *oscilaţiilor în timp* a variabilelor statistice, cu ajutorul *seriilor în timp — serii dinamice*, cum li se mai spune.

O serie dinamică oglindește, printr-un șir de indicatori numerici, procesul de schimbare și dezvoltare a unei colectivități statistice date, la intervale de timp succesive.

Trebuie să subliniem că în concepția statisticii socialiste seriile în timp au o trăsătură caracteristică esențială; ele sînt construite numai pentru colectivități *omogene*. Aceasta înseamnă că se poate vorbi de schimbarea și dezvoltarea numerică — deci cantitativă — a unei *colectivități date*, atîta timp cît *esența ei calitativă rămîne aceeași*. În alte condiții, procesul de cunoaștere statistică cu ajutorul seriilor în timp nu mai are caracterul unui proces dialectic de cunoaștere științifică.

Esența, calitatea fenomenelor și proceselor oglindite în seriile dinamice este stabilită pe baza unei *analize concrete* a datelor statistice, ținîndu-se seama de specificul fenomenelor și proceselor social-economice studiate și de acțiunea legilor obiective care le guvernează.

Condiția omogenității seriilor dinamice nu exclude posibilitatea ca într-un tabel statistic să figureze date ce caracterizează perioade calitativ deosebite. Chiar în publicațiile noastre statistice, seriile întocmite pe perioade mai mari includ, de exemplu, și date pe anul 1938 cu care se compară datele din perioada desăvîrșirii construcției socialiste în țara noastră. Ceea ce este esențial în elaborarea seriilor dinamice este precizarea exactă a problemei pe care o supunem studiului. Dacă ne propunem să scoatem în evidență tocmai două dezvoltări de tip diferit pentru a releva deosebirea dintre ele, atunci, evident, vom folosi date ce se referă la perioade cu legi deosebite de dezvoltare. Dacă ne propunem însă studiul dezvoltării unui fenomen de tip dat, nu ne este îngăduit să folosim decît date calitativ omogene, caracterizînd tipul respectiv.

Mai trebuie să menționăm că pentru a asigura seriilor dinamice o deplină omogenitate, la întocmirea lor trebuie să eliminăm toate viciile care afectează de obicei comparabilitatea indicatorilor ce formează variantele seriei. Trebuie asigurată comparabilitatea din punctul de vedere al volumului colectivității studiate, al definirii unității de observare, al spațiului la care se referă datele, al prețurilor folosite etc. Nu vom insista asupra acestor probleme care sînt expuse pe larg în cursurile de statistică teoretică.

O altă trăsătură caracteristică unei serii dinamice omogene constă în faptul că fenomenele observate și oglindite în indicatorii care formează seria nu sînt fenomene independente. Acest lucru este desigur firesc pentru observarea schimbării în timp a unuia și aceluiași fenomen, avînd aceeași esență.

Dacă, de pildă, la o întreprindere siderurgică s-a produs în cursul lunii ianuarie o anumită cantitate de oțel, exprimată în tone, nivelul acestui indicator nu poate fi considerat drept un fenomen independent, față de nivelul lui din luna decembrie a anului precedent, atîta timp cît de la o perioadă la alta nu s-au schimbat condițiile tehnico-organizatorice ale producției. Acest exemplu contrastează puternic cu exemplul aruncării succesive a unei monede în care rezultatele unei experiențe nu influențează rezultatele experienței următoare.

Devine clar, deci, că schemele teoretice ale calculului probabilităților bazate pe independența fenomenelor nu sînt în general aplicabile

Scanned with OKEN Scanner

seriilor dinamice ce se întocmesc pentru cunoaşterea fenomenelor social-economice. În acest caz se aplică *teoria proceselor stohastice.*

În ultimele trei decenii s-a creat în calculul probabilităţilor teoria funcţiilor aleatoare. Fie $y = f(x)$ o funcţie obişnuită de x. Aceasta înseamnă că între mulţimea valorilor luate de variabila x şi mulţimea valorilor luate de funcţia y am creat o corespondenţă, astfel că fiecărui element x să-i corespundă un element y. Dacă elementele y sînt variabile aleatoare vom spune că $y = f(x)$ este o funcţie aleatoare. Un proces stohastic este o funcţie aleatoare în care variabila x ia valori reale, putînd fi considerate ca diferite momente în timp. Ca exemplu de proces stohastic vom da o succesiune de experienţe consecutive, executate în modul următor: prima experienţă constă din extragerea unei bile dintr-o urnă, conţinînd 10 bile albe şi 10 bile roşii; a doua experienţă dintr-o extracţie din aceeaşi urnă, în care n-am înapoiat bila extrasă din prima experienţă şi aşa mai departe. Recunoaştem schema bilei nerevenite sau schema hipergeometrică de care ne-am ocupat pe larg. Variabila x reprezintă rangul experienţei. x poate lua prin urmare valorile 1, 2, 3, ..., 20. Pentru fiecare x avem o variabilă aleatoare corespunzătoare. Dacă $x = 1$, avem variabila aleatoare corespunzătoare primei experienţe, constînd din extracţia unei bile albe cu probabilitatea $\frac{1}{2}$ şi a unei bile roşii cu probabilitatea $\frac{1}{2}$. Pentru $x = 2$, avem experienţa a doua. Dacă în prima extracţie am scos o bilă albă, atunci pentru $x = 2$ corespunde o variabilă aleatoare constînd din extracţia unei bile albe cu probabilitatea $\frac{9}{19}$ şi a unei bile roşii cu probabilitatea $\frac{10}{19}$.

Dacă bila extrasă din prima experienţă este roşie, atunci pentru $x=2$ corespunde variabila aleatoare constînd din extracţia unei bile albe cu probabilitatea $\frac{10}{19}$ şi a unei bile roşii cu probabilitatea $\frac{9}{19}$. În acelaşi mod, se găsesc variabilele aleatoare, care corespund lui $x = 3,4, \ldots, 20$.

Se vede acum diferenţa care există între o variabilă aleatoare şi între un proces stohastic.

La o variabilă aleatoare obţinem o valoare reală cu o probabilitate determinată. La un proces stohastic obţinem o variabilă aleatoare cu o probabilitate determinată.

Să notăm prin $\begin{pmatrix} 1 & 0 \\ p & q \end{pmatrix}$ variabila aleatoare constînd din extragerea unei bile albe cu probabilitatea p şi neextragerea unei bile albe (extragerea unei bile roşii) cu probabilitatea q. La variabila aleatoare din cazul de faţă, putem obţine valoarea 1 cu probabilitatea p şi valoarea zero cu probabilitatea q.

Să considerăm în cazul procesului stohastic de mai sus, pe care l-am dat ca exemplu, $x = 2$. În acest caz cu probabilitatea $\frac{1}{2}$ vom avea variabila aleatoare $\begin{pmatrix} 1 & 0 \\ \frac{10}{19} & \frac{9}{19} \end{pmatrix}$ şi cu probabilitatea $\frac{1}{2}$ variabila aleatoare $\begin{pmatrix} 1 & 0 \\ \frac{9}{19} & \frac{10}{19} \end{pmatrix}$.

Pentru $x = 3$, cu probabilitatea $\frac{9}{38}$ vom avea variabila aleatoare $\begin{pmatrix} 1 & 0 \\ \frac{8}{18} & \frac{10}{18} \end{pmatrix}$ şi cu probabilitatea $\frac{20}{38}$ variabila aleatoare $\begin{pmatrix} 1 & 0 \\ \frac{9}{18} & \frac{9}{18} \end{pmatrix}$ şi cu probabilitatea $\frac{9}{38}$ variabila aleatoare $\begin{pmatrix} 1 & 0 \\ \frac{10}{18} & \frac{8}{18} \end{pmatrix}$.

Prima variabilă aleatoare corespunde cazului cînd în primele două extracţii au ieşit două bile roşii, a doua cazului cînd în primele două extracţii au ieşit o bilă roşie şi una albă, iar ultima cazului cînd în primele două extracţii au ieşit două bile albe.

Cunoaşterea ştiinţifică a dezvoltării fenomenelor în timp constituie, aşadar, o problemă complexă. Acest lucru este explicabil prin faptul că schimbările care au loc în dezvoltarea unei colectivităţi studiate se referă la toate elementele ei : se schimbă, în primul rînd, *volumul* şi *structura* colectivităţii, iar în al doilea rînd, se schimbă şi *caracteristicile* proprii diferitelor unităţi ale colectivităţii, *forma* şi *caracterul* legăturii dintre aceste caracteristici. Pe de altă parte, toate aceste schimbări, la rîndul lor, sînt interdependente.

Continua mişcare şi schimbare a fenomenelor vieţii social-economice, oglindite în seriile economice, sînt în funcţie de condiţiile concrete, reflectînd rezultatul acţiunii unui număr foarte mare de factori.

Influenţa acţiunii acestor factori este interdependentă şi imprimă dezvoltării fenomenelor şi proceselor social-economice diferite oscilaţii (variaţii) şi, totodată, le imprimă şi o anumită regularitate, legitate.

Diversele moduri de manifestare a dezvoltării fenomenelor se pot disocia, pentru a fi mai bine studiate. O asemenea disociere are desigur un caracter pur teoretic.

Oscilaţiile şi schimbările care pot modifica dezvoltarea unui fenomen sînt de diferite ordine, corespunzînd efectului acţiunii unor anumiţi factori

Astfel, în primul rînd, o serie în timp, reprezentată grafic, poate să ne scoată în evidenţă o anumită *tendinţă centrală* *) în dezvoltarea fenomenului, o direcţie fundamentală de mişcare.

Noţiunea de tendinţă centrală este deosebit de utilă procedeelor de cercetare ştiinţifică a seriilor în timp ; ea poate ilustra o mişcare ce se dezvoltă continuu în sens crescător, descrescător sau constant (foarte rar la fenomene social-economice). Schimbările particulare şi întîmplătoare nu sînt compatibile cu ideia de tendinţă centrală.

În al doilea rînd, într-o serie în timp se pot observa aşa-numitele *oscilaţii sezonale* (periodice) care au loc prin efectul acţiunii unor factori naturali, de ordin social sau din cauza unor neregularităţi calendaristice. De pildă, mişcarea mărfurilor pe căi ferate, desfacerile bunurilor de larg consum etc. sînt supuse oscilaţiilor sezonale ce se repetă periodic, cu mici abateri, uneori cu schimbări progresive foarte mici. Aceste oscilaţii au, în general, o periodicitate scurtă.

În al treilea rînd, se observă uneori *oscilaţii incidentale* (întîmplătoare) care apar ca rezultat al acţiunii unor factori întîmplători cum ar fi, de pildă, catastrofe din cauze naturale (cutremure, inundaţii etc).

Aceste oscilaţii sînt denumite de către unii cercetători şi *oscilaţii reziduale*. Influenţa factorilor întîmplători se manifestă în schimbarea

*) În terminologia statistică internaţională a căpătat circulaţie expresia „*trend*" din limba engleză, mai ales în ţările cu limba engleză şi germană. Unii autori folosesc şi expresiile „*tendinţă seculară*" sau „*tendinţă de lungă durată*", deşi sînt departe de a oglindi noţiunea ce o exprimă.

diferitelor variante ale variabilei în timp, modificînd mişcarea de tendinţă centrală ca şi oscilaţiile sezoniere.

În rîndul factorilor întîmplători trebuie să trecem şi factorii de mică importanţă, de care în general nu ţinem sau nu putem ţine seama şi care exercită prin însumarea lor anumite influenţe asupra dezvoltării seriei în timp.

În seriile dinamice care caracterizează fenomenele social-economice din ţările capitaliste se manifestă şi aşa numitele *oscilaţii ciclice*, determinate de periodicitatea succesiunii ciclurilor economice în procesul reproducţiei din economia capitalistă. Prin însăşi definiţia lor, ciclurile se repetă în mod aproximativ la acelaşi interval de timp ceea ce se reflectă şi în mişcarea indicatorilor din seriile dinamice respective.

Am arătat mai sus că pentru efectuarea cercetării ştiinţifice unele forme ale dezvoltării fenomenului pot fi disociate. Se disociază, de obicei oscilaţiile de un anumit tip (tendinţa centrală, oscilaţiile sezonale etc.) pe care vrem să le cercetăm separat de celelalte oscilaţii ce se oglindesc în seria dinamică luată în studiu.

Problema centrală care se pune de obicei cu prilejul disocierii diferitelor oscilaţii este aceea de a releva în ce mod elementele fundamentale ale mişcării fenomenului, asociindu-se între ele, au dat naştere seriei dinamice reale. Pentru descompunerea seriei dinamice pot fi folosite diferite procedee, ele însă trebuie să se întemeieze pe analiza calitativă a fenomenelor social-economice oglindite în serie.

Aşadar, disocierea, descompunerea unei dezvoltări în componentele arătate prezintă un deosebit interes teoretic, ea fiind în măsură să pună statistic în evidenţă interdependenţa dintre două sau mai multe fenomene social-economice. Ea prezintă totodată un interes practic necontestat. Cunoscînd, de pildă, variaţiile sezonale ale unui fenomen care, în condiţiile desfăşurării lui normale poate afecta considerabil activitatea ritmică a unei întreprinderi sau organizaţii economice, organele de conducere planificată a sectorului respectiv de activitate vor putea lua măsuri organizatorice corespunzătoare, înlăturînd efectele dăunătoare ale sezonalităţii asupra realizării economiilor, rentabilităţii etc.

2. Clasificarea seriilor dinamice şi reprezentarea lor grafică

Datele ce formează o serie dinamică pot avea diferite semnificaţii. În funcţie de semnificaţia datelor, seriile dinamice se împart în *serii de momente* — după denumirea folosită de unii autori — *serii integrale* şi *serii de intervale* sau *serii diferenţiale*.

Se numesc *serii de momente* acele serii care caracterizează fenomenul în diferite momente succesive ale timpului.

O serie de momente poate fi ilustrată prin exemplul din tabelul 41.

O particularitate caracteristică a seriilor de momente constă în aceea că indicatorii succesivi ai acestor serii nu se pot însuma ; o asemenea însumare n-ar avea nici un înţeles.

Scanned with OKEN Scanner

Tabelul 41

Numărul mediu scriptic al muncitorilor în industria socialistă din R.P.R. pe anii 1950—1958 *)

Anii	1950	1951	1952	1953	1954	1955	1956	1957	1958	1959
Numărul mediu scriptic al muncitorilor (în mii oameni)	640,4	728,2	813,4	865,4	895,2	916,4	936,1	938,0	979,2	1013,9

*) Datele extrase din *Anuarul Statistic al R.P.R.*, 1959, p. 125 și 1960, p. 171.

Seriile de intervale (diferențiale) sînt serii ai căror indicatori caracterizează volumul fenomenului pe intervale de timp determinate.

Exemplu de serie de intervale este dat în tabelul 42.

Datele acestui tabel ilustrează cum s-a schimbat volumul anual al producției de energie electrică în perioada 1950—1958.

Dacă ne referim, de pildă, la anul 1950, volumul producției de energie electrică în acest an (în perioada de la 1 ianuarie 1950 pînă la 1 ianuarie

Tabelul 42

Producția de energie electrică în R.P.R. în anii 1950—1958 *)

Anii	Producția de energie electrică (milioane kWh)
1950	2 113
1951	2 473
1952	2 886
1953	3 411
1954	3 697
1955	4 340
1956	4 930
1957	5 440
1958	6 184
1959	6 824

*) Date extrase din *Anuarul Statistic al R.P.R., 1960*, p. 168.

1951) a atins un nivel de 2,113 miliarde kWh; momentul inițial care marchează producția de energie electrică este dată de 1 ianuarie 1950. Pentru anul 1951, în care s-a produs un volum mai mare de energie electrică, momentul inițial de la care s-a făcut calculul producției este 1.I.1951, iar momentul final, 1.I.1952.

Este evident că datele unei asemenea serii dinamice se pot însuma și această însumare are un sens economic bine determinat.

Producția de energie electrică pe perioada 1950—1952 este de 4,586 miliarde kWh (2,113 + 2,473); pe perioada 1950—1953, de 7 472 miliarde kWh ș.a.m.d. Aceste date reprezintă volumul cumulat al producției de energie electrică, începînd cu un anumit moment.

Scanned with OKEN Scanner

Datele seriei de intervale se referă, deci, la anumite *perioade* de timp, exprimînd *rezultatele* dezvoltării, în aceste perioade, a unui proces sau fenomen bine determinat.

Datele care formează o serie dinamică pot fi exprimate în mărimi absolute sau în mărimi medii sau orice alte valori tipice. În funcţie de forma indicatorilor statistici folosiţi la construirea seriilor dinamice putem face subîmpărţiri ale acestora.

În tabelul 43 dăm un exemplu de serie dinamică în care datele pe diferiţi ani caracterizează consumul anual mediu de energie electrică pe un muncitor industrial.

Tabelul 43

Dezvoltarea consumului anual de energie electrică în kWh pe muncitor industrial (1950—1959) *)

1950	1951	1952	1953	1954	1955	1956	1957	1958	1959
2 104	2 188	2 296	2 586	2 665	3 089	3 332	3 626	3 859	4 105

*) Sursa: Revista de Statistică, 4,6, (1960) (după articolul publicat de L. Rădulescu şi D. Botoşăneanu).

Dezvoltarea industriei şi extinderea folosirii energiei electrice în activitatea industrială au fost cauza creşterii mereu mai mari a consumului de energie ce a revenit unui muncitor industrial în perioada considerată.

În lucrările sale, statistica socialistă foloseşte toate tipurile de serii dinamice arătate. Mai trebuie arătat că adeseori în practică se folosesc şi serii dinamice *complexe* în care apar mai multe date cronologice paralele între care există o strînsă legătură de *sistem*.

Pentru reprezentarea *grafică* a seriei de intervale se foloseşte de obicei *historiograma*, un grafic analog histogramei, utilizat la reprezentarea distribuţiilor de frecvenţă : pe axa absciselor se notează intervalele de timp corespunzătoare dinamicii reprezentate, iar pe axa ordonatelor, datele care ilustrează conţinutul seriei dinamice. Indicatorii seriei de intervale sînt reprezentaţi prin dreptunghiuri dispuse unul lîngă celălalt. Pentru ilustrare dăm mai jos (vezi fig. 29) reprezentarea grafică a seriei de intervale din tabelul 42, privind producţia de energie electrică în R.P.R. în anii 1950—1958.

Pentru reprezentarea grafică a seriei dinamice ilustrată în tabelul 42 ar fi fost suficient să considerăm numai liniile continue de pe grafic; au fost trasate însă şi liniile punctate pentru explicaţii de detaliu.

Este uşor de văzut că un asemenea grafic ne ilustrează sugestiv nu numai volumul anual al producţiei de energie electrică, ci şi sporurile anuale ale acestui volum, raportul ce se formează între nivelurile vecine ale volumului producţiei. Diagrama liniară ne sugerează ideea că nivelurile anuale ale volumului producţiei de energie electrică nu sînt valori izolate, independente, ci condiţionate unul de altul. Diagrama punctată pune în relief schimbarea nivelurilor anuale ale producţiei de energie electrică.

Graficul din figura 29 poate servi şi pentru a determina volumul producţiei de energie electrică „prin interpolare", considerînd anumite

Scanned with OKEN Scanner

perioade de timp pentru care nu avem date. A se vedea producția de energie electrică în perioada 1.VII.1957—1.VII.1958, reprezentată pe diagramă printr-un dreptunghi care atinge pe axa ordonatelor o înălțime de

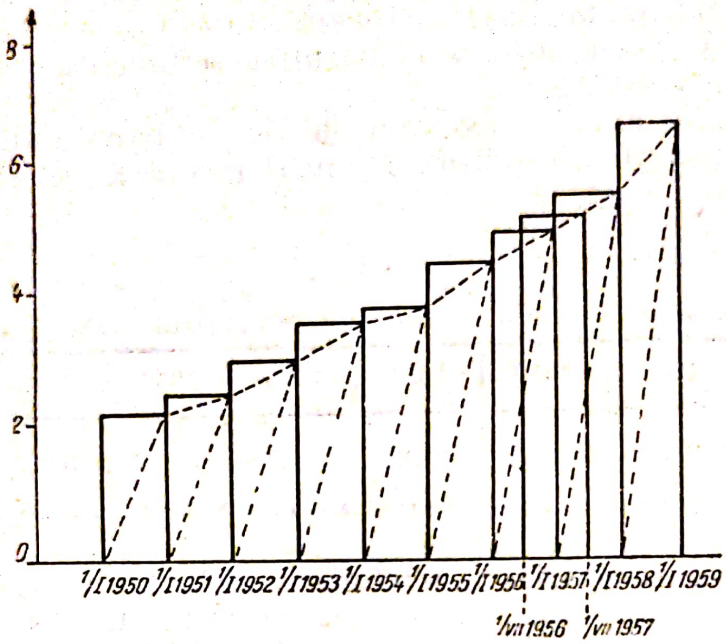


Fig. 29

aproximativ 5,16 miliarde kWh. Trebuie să atragem atenția asupra preciziei relative a unei asemenea „interpolări".

Reprezentarea grafică a seriei dinamice de momente este mai simplă decît a seriei de intervale. Ea se face cu ajutorul unei linii frînte care unește vîrfurile ordonatelor reprezentînd indicatorii seriei dinamice, duse perpendicular pe axa absciselor în punctele de momente ale timpului.

Pentru ilustrare, în figura 30 sînt reprezentate datele din tabelul 41 reprezentînd numărul mediu scriptic al muncitorilor din industria socialistă din R.P.R. pe anii 1950—1958.

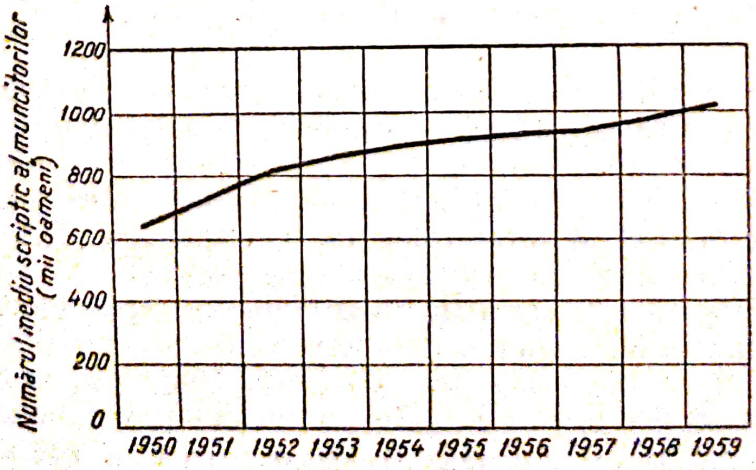


Fig. 30

Semnificaţia acestui grafic este clară. El oglindeşte dinamica numărului mediu scriptic al muncitorilor din industria socialistă a R.P.R. Întrucît este vorba de date medii anuale, punctele pe axa absciselor se iau la jumătatea intervalului reprezentînd perioada de un an.

Toate celelalte serii dinamice pot fi reprezentate grafic ca în figurile 29 şi 30.

O formă interesantă de grafic poate fi folosită la reprezentarea seriei dinamice din tabelul 43.

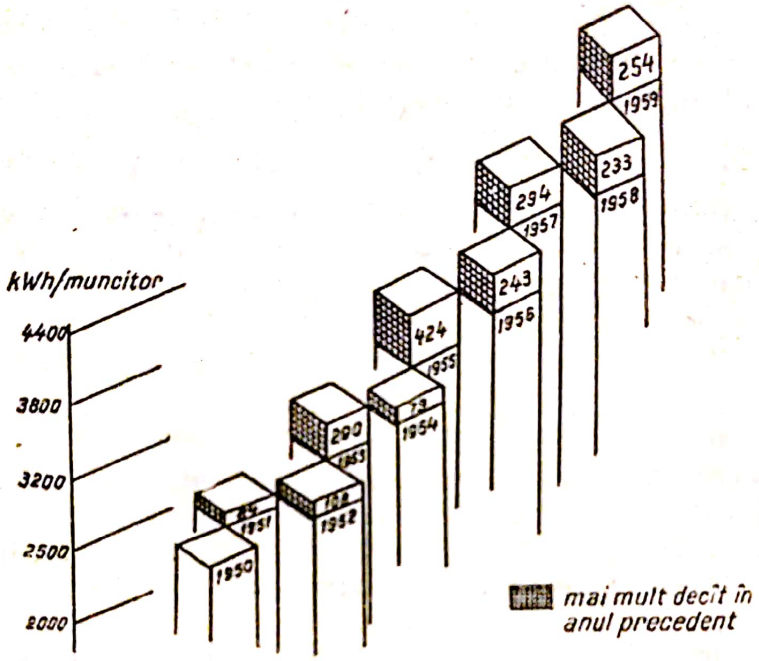


Fig. 31

3. Indicatorii seriei dinamice

Din cele expuse pînă acum, s-a putut reţine că o serie dinamică cuprinde în primul rînd mai mulţi *indicatori de nivel*, fiecare caracterizînd volumul fenomenului la diferite momente ale timpului sau pe anumite intervale de timp.

Relevarea schimbării fenomenului sau procesului oglindit într-o serie dinamică se poate face şi printr-o simplă comparare a doi indicatori succesivi ai seriei, dar acest lucru încă nu ne poate oferi o caracterizare suficientă pentru a trage concluzii asupra dezvoltării fenomenului sau procesului respectiv.

O analiză mai adîncă a unei serii dinamice necesită calculul mai multor indicatori derivaţi, sintetici.

Este vorba de calculul *sporului absolut*, care ne arată cu cîte unităţi în expresie absolută, fiecare nivel succesiv al seriei este mai mare sau mai mic decît nivelul precedent, diferenţa făcîndu-se algebric. Acest indicator caracterizează, într-o măsură oarecare, intensitatea dezvoltării fenomenului.

Dacă notăm cu

$$y_1, y_2, \ldots, y_n$$

indicatorii de nivel ai unei serii dinamice şi cu

$$\delta_{2/1}, \delta_{3/2}, \delta_{4/3}, \ldots, \delta_{n/n-1},$$

sporurile absolute, vom avea

$$\delta_{2/1} = y_2 - y_1; \quad \delta_{3/2} = y_3 - y_2; \quad \ldots; \quad \delta_{r/n-1} = y_n - y_{n-1}$$

valorile acestor sporuri.

Se poate uşor observa că suma sporurilor absolute ale seriei dinamice ne dă sporul absolut al întregii serii :

$$\delta_{n/1} = y_2 - y_1 + y_3 - y_2 + y_4 - y_3 + \ldots + y_n - y_{n-1} = y_n - y_1. \quad \text{(XII, 3.1)}$$

Pe baza sporului absolut al întregii serii se poate calcula *sporul mediu absolut*, denumit de unii autori şi *viteza medie* de schimbare pentru seria dată ; notînd cu v acest indicator, avem

$$v = \frac{\delta_{n/1}}{n-1} = \frac{y_n - y_1}{n-1}, \quad \text{(XII, 3.2)}$$

în care n este numărul termenilor seriei.

Intensitatea dezvoltării fenomenului poate fi însă caracterizată şi mai puternic prin calculul unor indicatori relativi, raportîndu-se indicatorii de nivel ai seriei la un alt indicator din aceeaşi serie, luat ca bază de comparaţie. Asemenea indicatori relativi sînt *ritmul de creştere* şi *ritmul sporului*.

Calculul acestor indicatori nu implică nici un fel de dificultăţi cu caracter matematic.

Folosind aceeaşi notaţie şi introducînd simbolul i pentru indicatorii individuali (coeficienţi) ai ritmului de creştere, avem

$$i_{2/1} = \frac{y_2}{y_1}; \quad i_{3/2} = \frac{y_3}{y_2}; \quad \ldots; \quad i_{n/n-1} = \frac{y_n}{y_{n-1}}.$$

Produsul lor ne dă $i_{n/1}$, după cum se poate uşor demonstra

$$\frac{y_2}{y_1} \cdot \frac{y_3}{y_2} \cdot \frac{y_4}{y_3} \ldots \frac{y_n}{y_{n-1}} = \frac{y_n}{y_1} = i_{n/1}. \quad \text{(XII, 3.3)}$$

Notînd, în sfîrşit, prin k *ritmul sporului* avem

$$k = \frac{y_n - y_{n-1}}{y_{n-1}}. \quad \text{(XII, 3.4)}$$

După cum se vede, între cei doi indicatori relativi există următoarea relaţie :

$$k = \frac{y_n - y_{n-1}}{y_{n-1}} = i_{n/n-1} - 1. \quad \text{(XII, 3.5)}$$

Scanned with OKEN Scanner

Indicatorii ce caracterizează ritmul creşterii şi ritmul sporului pot fi exprimaţi în procente, formă sub care se întîlnesc adeseori în practică.

Probleme mai interesante se ridică la calculul *ritmului mediu de creştere* a unui fenomen social-economic. Asupra acestor probleme vom reveni în capitolul XIV după ce vom expune în prealabil procedeele de ajustare şi interpolare a seriilor dinamice.

Un alt indicator care ne poate caracteriza sintetic nivelul seriei dinamice este *nivelul mediu* al seriei.

Pentru calculul nivelului mediu al unei serii dinamice de intervale se foloseşte media aritmetică

$$\bar{y} = \frac{y_1 + y_2 + \ldots + y_n}{n}. \qquad \text{(XII, 3.6)}$$

Pentru seriile dinamice de momente, formula generală folosită este aceea a mediei cronologice şi anume:

$$\bar{y}_{\text{cron.}} = \frac{y_1 + 2y_2 + 2y_3 + \ldots + 2y_{n-1} + y_n}{2(n-1)}. \qquad \text{(XII, 3.7)}$$

Demonstraţia acestei formule generale este simplă.

Să notăm cu $\frac{y_1 + y_2}{2}$, $\frac{y_2 + y_3}{2}$, $\ldots$, $\frac{y_n + y_{n-1}}{2}$, mediile grupelor de cîte doi indicatori de nivel. Se vede că numărul acestor medii de grupe este egal cu $n - 1$ şi făcînd media lor, obţinem media cronologică generală a seriei, adică

$$\left[\frac{y_1 + y_2}{2} + \frac{y_2 + y_3}{2} + \ldots + \frac{y_{n-1} + y_n}{2}\right] : (n-1) =$$

$$= \left[\frac{y_1}{2} + \frac{y_2}{2} + \frac{y_2}{2} + \frac{y_3}{2} + \frac{y_3}{2} + \ldots + \frac{y_{n-1}}{2} + \frac{y_n}{2}\right] : (n-1) =$$

$$= \frac{\frac{y_1}{2} + \frac{2y_2}{2} + \frac{2y_3}{2} + \ldots + \frac{2y_{n-1}}{2} + \frac{y_n}{2}}{n-1} =$$

$$= \frac{y_1 + 2y_2 + \ldots + 2y_{n-1} + y_n}{2(n-1)} = \frac{\frac{y_1}{2} + y_2 + \ldots + y_{n-1} + \frac{y_n}{2}}{n-1}.$$

Am obţinut formula identică cu (XII, 3.7).

Trebuie să observăm că formula generală a mediei cronologice (XII, 3.7) este aplicabilă numai *seriilor dinamice cu intervale egale.*

În cazul cînd avem intervale inegale, folosim formula mediei cronologice ponderate, ponderile fiind duratele de timp ale intervalelor inegale.

Scanned with OKEN Scanner

Dacă notăm termenii seriei cu y_i iar intervalele inegale ce separă termenii cu t_i, formula generală a mediei cronologice ponderate este:

$$\bar{x}_c = \frac{\left(y_1 \frac{t_1}{2}\right) + \left(y_2 \frac{t_1 + t_2}{2}\right) + \ldots + \left(y_n \frac{t_{n-1}}{2}\right)}{\frac{t_1}{2} + \frac{t_1 + t_2}{2} + \ldots + \frac{t_{n-1}}{2}}.$$

La stabilirea acestei formule vom ajunge examinînd mai adînc caracterul seriilor dinamice de momente.

Să considerăm o serie dinamică de momente

$$\begin{pmatrix} y_1 & y_2 \cdots y_n \\ t_1 & t_2 \ldots t_n \end{pmatrix}.$$

Aceasta înseamnă că la momentele de timp respective $t_1, t_2, \ldots, t_n$ seria dinamică a luat valorile respective $y_1, y_2, \ldots, y_n$.

$y_1, y_2, \ldots, y_n$ pot să reprezinte de exemplu temperatura atmosferii în timpul unei zile luată la orele 0, 1, 2, ..., 24. În acest caz, $n = 25$ şi momentele $t_1, t_2, \ldots, t_n$ sînt respectiv egale cu 0, 1, 2, ..., 24. Pe baza acestor date ne propunem să stabilim temperatura medie a zilei.

Vom face în prealabil observaţia că temperatura putea fi luată nu numai la momentele t_i, $(i = 1, 2, \ldots, 25)$, ci la orice moment al zilei considerate şi este evident că cu cît am fi făcut mai multe observaţii ale temperaturii în perioada celor 24 de ore, cu atît temperatura medie a zilei ar fi fost mai exactă.

Să notăm pe axa Ot timpul şi pe axa Oy temperaturile

Reprezentată grafic, temperatura, care este o funcţie $y = f(t)$ de momentul considerat, să presupunem că are forma din figura 32.

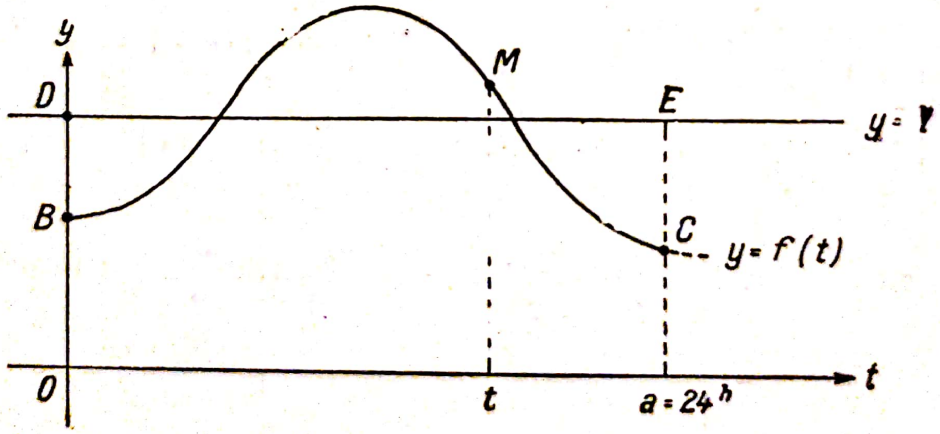


Fig. 32

La orice moment t, din timpul celor 24 de ore, există o temperatură determinată, egală pe figură cu ordonata Mt. Fie Y temperatura medie a zilei. Aria $OBMCa$ reprezintă, într-un anumit sens, suma temperaturilor din timpul zilei. Aria $ODEa$ reprezintă suma temperaturilor dacă temperatura în tot timpul zilei a fost egală cu temperatura medie Y. Y nu poate fi temperatură medie decît dacă cele două arii $OBMCa$ şi $ODEa$, a căror semnificaţie am arătat-o, sînt egale.

Însă

$$\text{aria } OBMca = \int_0^a f(t)dt,$$

$$\text{aria } ODEa = a\,Y.$$

Trebuie să avem

$$a\,Y = \int_0^a f(t)dt.$$

Deducem

(I) $$Y = \frac{1}{a}\int_0^a f(t)dt.$$

Relaţia (I) ne dă formula exactă a temperaturii medii. Ea nu poate fi aplicată în practică pentru că nu putem folosi decît un număr finit de observaţii ale temperaturii.

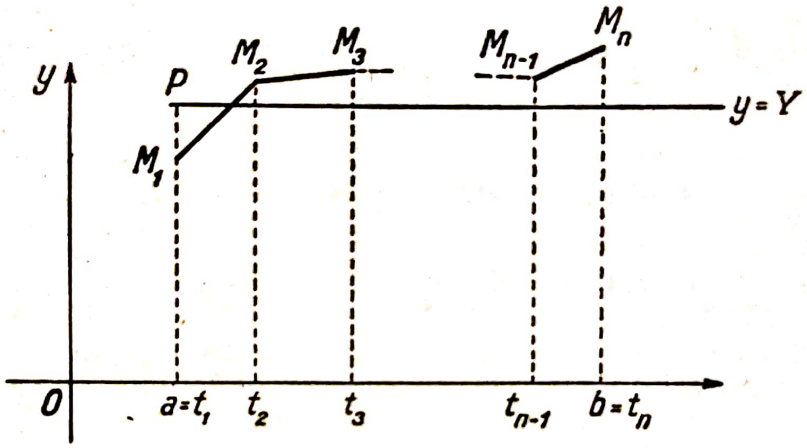


Fig. 33

Fie punctele $M_1(t_1, y_1)$, $M_2(t_2, y_2), \ldots, M_n(t_n, y_n)$ care corespund la n observaţii ale temperaturii făcute la momentele $t_1, t_2, \ldots, t_n$. Însemnăm

$$t_1 = t_2 - t_1, \quad t_2 = t_3 - t_2, \ldots, t_{n-1} = t_n - t_{n-1}.$$

Aria $OBMCa$ din figura 32, este aproximată în cazul de faţă prin aria $aM_1M_2 \ldots M_nb$. Ea este formată dintr-o sumă de trapeze

$$aM_1M_2t_2,\ t_2M_2M_3t_3, \ldots, t_{n-1}M_{n-1}M_nt_n$$

avînd respectiv ariile

$$\frac{y_1 + y_2}{2}t_1, \quad \frac{y_2 + y_3}{2}t_2, \ldots, \frac{y_{n-1} + y_n}{2}t_{n-1}.$$

Suma acestor arii este

$$y_1\frac{t_1}{2} + y_2\frac{t_1 + t_2}{2} + \ldots + y_n\frac{t_{n-1}}{2}.$$

Dacă toate temperaturile y_i, $(i = 1, 2, \ldots, n)$, ar fi egale cu temperatura medie Y, suma de mai sus ar fi:

$$Y\frac{t_1}{2} + Y\frac{t_1 + t_2}{2} + \ldots + Y\frac{t_{n-1}}{2}.$$

Egalînd aceste două arii rezultă

$$Y\left[\frac{t_1}{2}+\frac{t_1+t_2}{2}+\ldots+\frac{t_{n-1}}{2}\right]=y_1\frac{t_1}{2}+y_2\frac{t_1+t_2}{2}+\ldots+y_n\frac{t_{n-1}}{2}$$

sau

$$Y=\frac{y_1\frac{t_1}{2}+y_2\frac{t_1+t_2}{2}+\ldots+y_n\frac{t_{n-1}}{2}}{\frac{t_1}{2}+\frac{t_1+t_2}{2}+\ldots+\frac{t_{n-1}}{2}},$$

adică tocmai formula generală a mediei cronologice pe care am vrut s-o găsim.

Scanned with OKEN Scanner

Scanned with OKEN Scanner

Capitolul XIII

Ajustarea şi interpolarea seriilor dinamice

1. Procedee elementare de ajustare

Sarcina principală la întocmirea unei serii dinamice am văzut că este aceea de a pune în evidenţă o legitate ce se manifestă în dezvoltarea fenomenului studiat, oglindit în datele seriei. Sînt numeroase cazuri cînd datele reale ale seriei scot în evidenţă cu destulă claritate legitatea dezvoltării, tendinţa ei centrală. De exemplu, dacă examinăm seria datelor din tabelul 44 şi historiograma corespunzătoare (fig. 35) constatăm cu uşurinţă existenţa unei dezvoltări nete a populaţiei oraşului Bucureşti de-a lungul timpului care poate fi relevată şi printr-un procedeu de *ajustare*.

În legătură cu folosirea termenilor de „ajustare" şi „interpolare", credem că sînt utile unele precizări prealabile.

După unii autori, rolul pe care îl joacă curba tendinţei centrale faţă de historiograma efectivă se poate identifica — din punctul de vedere strict formal — cu cel al curbei de repartiţie, exprimînd o anumită lege de probabilitate, faţă de poligonul frecvenţelor empirice. Într-adevăr, nici un cercetător n-ar încerca să ajusteze cu o altă curbă decît cea binomială (or cu cea normală în cazul unui mare număr de cazuri observate) pentru cazul fenomenelor asemănătoare din punct de vedere structural cu urna lui Bernoulli cu două stări. În timp ce pentru forma curbelor exprimînd anumite legi teoretice de probabilitate avem unele indicaţii aprioriice, pentru tendinţa centrală de lungă durată a aceluiaşi fenomen putem adopta formule din cele mai variate din punct de vedere analitic.

Există, prin urmare, spun aceşti autori, o diferenţă fundamentală între bazele operaţiunii de ajustare şi ale celei de interpolare, rezervînd folosirea termenului de *interpolare* îndeosebi pentru analiza seriilor în timp, la determinarea tendinţei centrale.

Credem că folosirea diferenţiată a celor doi termeni este o problemă atît de conţinut cît şi de formă a operaţiunilor pe care le efectuăm.

1°. În cele ce urmează vom considera că sîntem în cazul unei *probleme de interpolare* cînd — fiind cunoscut un şir de valori ale unei mărimi, corespunzînd unor abscise date — ni se pune problema *determinării valorii corespunzătoare unei alte abscise date*. Valorile cunoscute

variază uneori în mod statistic, alteori pot reprezenta valorile numerice ale unei funcţii, cunoscută sau necunoscută.

Procedeul clasic al interpolării constă în a face să treacă prin punctele corespunzătoare absciselor şi ordonatelor date o curbă a cărei ecuaţie ne este cunoscută. Ordonata punctului situat pe curbă, corespunzînd abscisei intermediare, ne dă apoi valoarea căutată.

Dacă se cunosc $n + 1$ valori

$$y_0, y_1, y_2, \ldots, y_n \tag{a}$$

pe care le ia o funcţie pentru $n + 1$ valori ale variabilei

$$x_0, x_1, x_2, \ldots, x_n \tag{b}$$

operaţia de interpolare constă, deci, în a determina valorile funcţiei pentru valori intermediare ale variabilei independente.

În general valorile (a) sînt date experimentale, prin urmare, funcţia ale cărei valori sînt aceste numere este, de obicei, necunoscută. Pusă în acest mod, problema interpolării este nedeterminată, deoarece există un număr infinit de funcţii care să ia valorile (a) pentru valorile (b) ale variabilei independente. Problema interpolării se poate rezolva înlocuind funcţia necunoscută printr-un polinom de gradul n şi în acest caz problema este complet determinată, deoarece există un singur polinom de gradul n care, pentru $n + 1$ valori ale variabilei independente, ia $n + 1$ valori date.

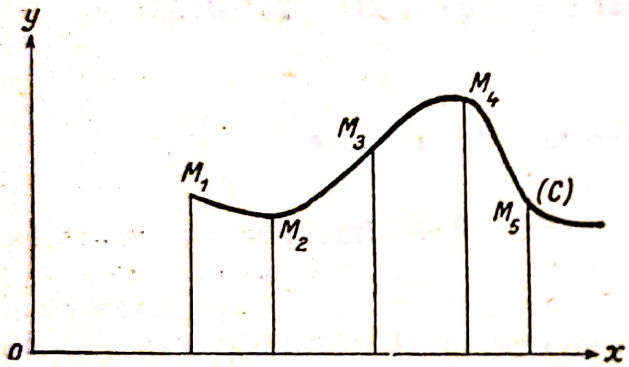


Fig. 34

Determinînd un asemenea polinom putem calcula valorile sale pentru valori intermediare oarecare ale lui x, aceste valori trebuie considerate, însă, ca nişte valori apropiate ale unei funcţii necunoscute.

Este evident că erorile comise sînt cu atît mai mici cu cît numărul valorilor date variabilei independente este mai mare şi deci, intervalele dintre ele sînt mai mici.

Interpolarea printr-un polinom se numeşte *interpolare parabolică*. Mai jos vom da cîteva formule generale de interpolare.

2°. Vom considera că sîntem în cazul unei probleme de *ajustare*, cînd ni se dă un şir de n valori:

$$y_1, y_2, \ldots, y_n$$

şi se cere să determinăm cu ajutorul lor un alt şir

$$y_1', y_2', \ldots, y_n'$$

care să reprezinte cît mai bine rezultatele cercetărilor experimentale.

Traducerea geometrică a enunţului problemei ne va lămuri mai bine.

Să considerăm n puncte în planul xOy

$$M_i (x_i, y_i)$$

unde $i = 1, 2, \ldots, n$. Ne propunem să trasăm o curbă C care să treacă printre aceste puncte (eventual şi prin unele dintre ele), dar care să fie cea mai apropiată de ele (vezi fig. 34).

Printre punctele desenate putem să tragem o infinitate de linii. O cale foarte simplă pentru determinarea celei mai convenabile linii constă în trasarea unei curbe pe care, în raport

cu datele noastre, o apreciem că se apropie cel mai mult de valorile $y_1, y_2, \ldots, y_n$. Asemenea ajustări grafice lasă un mare grad de libertate celui care efectuează ajustarea.

Un alt procedeu care să ducă la concluzii independente de voinţa celui care prelucrează datele constă în aplicarea unor formule, de mai înainte stabilite, întregului şir de valori date $y_1, y_2, \ldots, y_n$. În acest mod obţinem aşa numitele *ajustări mecanice*. În statistica social-economică interesează însă îndeosebi *ajustările analitice* în care se pune problema generală a ajustării şi anume aceea de a exprima datele empirice — date care, în general, sînt afectate de erori de observare — printr-o formulă care să le reprezinte cît mai bine. Astfel de ajustări se numesc ajustări analitice.

Formula pe care o utilizăm, desigur, nu va exprima întreaga dezvoltare a fenomenului care se supune, în general, legilor social-economice obiective şi — în anumite condiţii — legităţilor statistice; ea va exprima doar anumite funcţii de care fenomenul cercetat se îndepărtează foarte puţin şi trebuie să îndeplinească o condiţie esenţială şi anume să conducă la calcule cît mai simple şi mai puţine.

Vom expune mai departe cîteva ajustări analitice din cele mai des folosite în aplicaţii.

Tabelul 44

Populaţia oraşului Bucureşti la recensămintele efectuate între anii 1831 – 1956

Anul recensămîntului	Total populaţie
1831	58 794
1859	121 734
1878	177 646
1889	184 488
1894	232 009
1899	282 071
1912	341 321
1930	639 040
1941	992 536
1948	1 041 807
1956	1 236 908

Notă: Datele se referă la teritoriul oraşului din anul respectiv. Din anul 1930, inclusiv comunele subordonate. (Vezi: *Anuarul Statistic al oraşului Bucureşti*, 1959, p. 36).

Printr-o abstractizare care nu-şi găseşte o deplină justificare logică, putem considera linia *dreaptă* (vezi fig. 35 linia dreaptă întreruptă) sau linia curbă care urmăreşte deaproape alura generală a historiogramei (vezi linia punctată) ca reprezentînd tendinţa centrală a dezvoltării populaţiei oraşului Bucureşti.

Este evident că ambele procedee sînt pur *descriptive*; ele apelează la intuiţia cercetătorului de a desprinde tendinţa centrală şi la îndemînarea lui de a o reprezenta. Este vorba, deci, de un procedeu de determinare vizuală grafică a tendinţei de bază a fenomenului, căruia îi sînt proprii impreciziile inerente procedeului însuşi.

Mai trebuie arătat că forma grafică a tendinţei centrale vizuale poate lua diferite înfăţişări; în afară de forma dreptei sau curbei exponenţiale (ca în fig. 35), ea mai poate lua forma unei linii frînte, forma unei

Scanned with OKEN Scanner

curbe parabolice sau hiperbolice sau chiar a unei drepte orizontale, care indică, în acest caz, lipsa de variație în timp.

Un alt procedeu elementar pentru ajustarea unei serii dinamice pentru a desprinde tendința centrală, este acela legat de folosirea indicatorului care caracterizează sporul mediu absolut al seriei considerate.

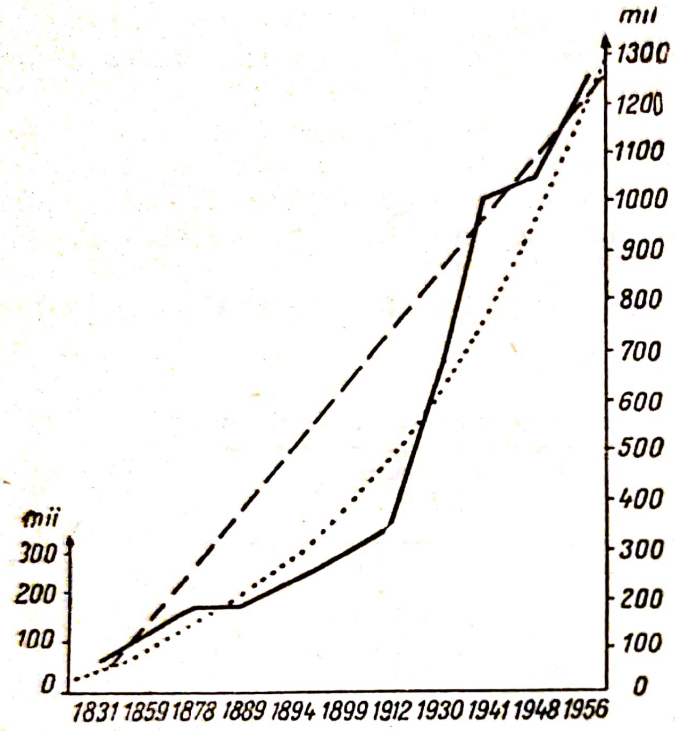


Fig. 35

De exemplu, considerînd datele din tabelul 45 ilustrînd dezvoltarea consumului anual de energie electrică în kWh pe un muncitor industrial și aplicînd formula (XIII,2.1) pentru calculul sporului mediu absolut, obținem un spor mediu al consumului de energie electrică pe un muncitor industrial egal cu 222,33 kWh.

Cunoscînd consumul de energie electrică în anul 1950 și sporul lui mediu anual, putem calcula un alt șir de indicatori pentru seria noastră care va reprezenta o *expresie statistică elementară* a dinamicii *medii* a consumului de energie electrică pe un muncitor industrial, în perioada 1950 — 1959 — o tendință centrală ca mijloc de exprimare sintetică a dinamicii medii a fenomenului nostru.

Concret, această serie ajustată pe baza sporului mediu absolut va avea următoarele valori (tabelul 45):

Tabelul 45

Consumul mediu anual de energie electrică în kWh pe muncitor industrial (1950—1959) (Ajustare pe baza sporului mediu anual al seriei efective)

Anii	Date ajustate pe baza sporului mediu anual $\left(r = \frac{y_n - y_1}{n-1}\right)$
1950	2 104,00
1951	2 326,33
1952	2 548,66
1953	2 770,99
1954	2 993,32
1955	3 215,65
1956	3 437,98
1957	3 882,31
1958	4 104,00*
1959	

*) Cu o aproximație în plus de 0,03, datorită periodicității fracției zecimale a nivelului sporului mediu anual.

Este evident că dacă aceste valori ajustate ale seriei dinamice vor fi reprezentate grafic, vom obține o linie dreaptă care va trece prin cele două variante extreme ale seriei și care sînt identice cu cele din seria dinamică efectivă.

Ecuația dreptei obținute prin ajustarea seriei efective, pe baza sporului mediu absolut, viteza ei medie de dezvoltare, este ecuația liniară:

$$Y_t = 222{,}33\ (t - 1) + Y_1$$

sau în formă generală

$$Y_t = \overline{V}\ (t - 1) + Y_1$$

în care prin t am notat anul de ordine în seria dinamică considerată.

2. Metoda mediilor mobile

Metoda mediilor mobile este o metodă de ajustare prin care — ca și la celelalte procedee elementare de ajustare — se înlocuiesc datele reale ale unei serii prin date medii, calculate pe baza unui număr determinat de termeni ai seriei reale; perioadele de timp pentru care se calculează mediile rămîn mereu egale, deplasîndu-se de la nivelul inițial al seriei dinamice cu cîte un interval, la calculul fiecărei medii, pînă la includerea în calcul și a nivelului final al seriei.

Metoda mediilor mobile se aplică îndeosebi la ajustarea seriilor cu oscilații mari, de obicei oscilații cu caracter sezonier, avînd drept obiectiv final relevarea tendinței centrale de dezvoltare a fenomenului cercetat.

Dacă există abateri întîmplătoare mari de la tendința centrală sau oscilații sezoniere pronunțate, făcînd media mai multor indicatori consecutivi ai seriei, abaterile sau oscilațiile într-un sens sau altul se compensează între ele, iar seria calculată va scoate în evidență tocmai tendința centrală de dezvoltare a fenomenului.

Numărul de termeni adoptat pentru calculul mediilor mobile poate fi ales arbitrar. De obicei se alege un număr impar de termeni. Analiza concretă a unei serii dinamice și aplicarea justă a metodei mediei mobile este însă incompatibilă cu alegerea arbitrară a acestui număr. Cercetătorul trebuie să fie atent și după o prealabilă analiză calitativă a fenomenului studiat el va alege acel număr de termeni ai seriei care să corespundă unor perioade de schimbări reale în evoluția fenomenului cercetat. El va trebui să țină seama, deci, de mărimea seriei, de frecvența și amplitudinea variațiilor și de alți factori ce acționează asupra fenomenului luat în studiu.

Este evident că dacă notăm termenii seriei cu $y_1, y_2, \ldots, y_n$ și alegem pentru calculul mediei mobile perioade de cîte trei termeni, mediile obținute $\bar{y}_I, \bar{y}_{II}, \ldots$ vor avea următoarea formă:

$$\bar{y}_I = \frac{y_1 + y_2 + y_3}{3};\quad \bar{y}_{II} = \frac{y_2 + y_3 + y_4}{3};\quad \bar{y}_{III} = \frac{y_3 + y_4 + y_5}{3} \ldots$$

$$\bar{y}_{n-1} = \frac{y_{n-2} + y_{n-1} + y_n}{3}.$$

După cum se vede, fiecare medie din trei termeni va cuprinde două elemente vechi şi unul nou. Datorită acestui fapt media calculată se deplasează de-a lungul întregii serii şi de aceea i s-a dat şi numele de medie mobilă.

În general, dacă numărul termenilor incluşi în calculul mediei este impar $(2n + 1)$, fiecare medie mobilă — indiferent dacă operăm asupra unei serii de momente sau serii de intervale — va înlocui termenul situat la mijlocul $(n + 1)$ fiecărei perioade supuse ajustării. Numărul acestor medii va fi egal cu $2n$, mai puţin decît numărul total al termenilor seriei.

Dacă în calculul mediei s-a inclus un număr par de termeni $(2n)$, atunci media se va plasa în punctul $n + \frac{1}{2}$ de pe intervalul luat în calcul, ceea ce înseamnă că seria de medii mobile nu va mai coincide cu termenii seriei efective. Practic, seria de medii mobile se poate ridica cu jumătate de interval mai sus aşa încît pentru calculul mediilor mobile se va putea folosi, de exemplu, următoarea formulă generală (pentru cazul cînd în calculul mediei se iau $2n + 1$ termeni):

$$\bar{y}_{\text{I}} = \frac{\frac{y_1}{2} + y_2 + y_3 + y_4 + \frac{y_5}{2}}{4} \text{ pentru prima medie}$$

$$\bar{y}_{\text{II}} = \frac{\frac{y_2}{2} + y_3 + y_4 + y_5 + \frac{y_6}{2}}{2} \text{ pentru a doua medie}$$

$$\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots$$

$$\bar{y}_{\text{XII}} = \frac{\frac{y_1}{2} + y_2 + \ldots + \frac{y_{13}}{2}}{12}, \text{ pentru ultima medie} \quad \text{(XIII, 2.1)}$$

în care $\bar{y}_k$ reprezintă o medie oarecare din lanţul mediilor mobile.

Pentru calculul indicilor sezonalităţii, fiecare termen al seriei reale se raportează la media generală a seriei, obţinîndu-se raportul

$$\frac{y_i}{\bar{y}_k}. \quad \text{(XIII, 2.2)}$$

Trebuie să atragem atenţia asupra unei probleme de importanţă deosebită.

Metoda mediei mobile este adeseori folosită de statisticieni. Orice statistician nu trebuie să uite însă că nu este vorba decît de un procedeu empiric care nu-i va permite să dezvăluie şi să formuleze o anumită legitate sau o lege oarecare. Legea nu poate fi stabilită decît pe baza unei analize materialist-dialectice a esenţei fenomenului studiat şi nu pe baza unei prelucrări aritmetice a seriei dinamice ce-l oglindeşte.

Din punct de vedere analitic, mediile tind doar să niveleze orice ondulaţie; acesta este principiul pe care se bazează întrebuinţarea lor.

Scanned with OKEN Scanner

Într-adevăr, după cum ştim, în statistică, orice medie reprezintă, sintetic, valoarea tipică a colectivităţii studiate. Or, indicatorii pe baza cărora se calculează mediile mobile nu pot fi asimilaţi indicatorilor ce se referă la unităţile unei colectivităţi în sensul ştiinţific al acestei categorii.

În consecinţă, mediile mobile îndeplinesc o funcţie „sui generis"; ele servesc drept procedeu de nivelare a ondulaţiilor, o simplă operaţie aritmetică care nu-şi găseşte o justificare teoretică riguroasă.

În continuare vom ilustra, prin exemple, folosirea metodei mediilor mobile.

Tabelul 46

Producţia de carne din R.P.R. (1958—1960) pe luni (Ajustarea prin metoda mediilor mobile)

— mii tone —

Denumirea lunilor	Anii			Media mobilă			Media lunară	Indicatori de sezon
	1958	1959	1960	1958	1959	1960		
Ianuarie	12,47	18,36	19,58	—	19,62	21,84	20,73	100,63
Februarie	14,21	14,97	18,48	—	19,58	21,98	20,78	100,87
Martie	13,34	14,96	23,32	—	19,40	22,27	20,93	101,60
Aprilie	18,78	19,07	19,77	—	19,36	22,53	20,94	101,65
Mai	16,48	16,77	19,65	—	19,60	22,73	21,16	102,71
Iunie	17,90	18,19	20,38	—	19,93	22,71	21,32	103,49
Iulie	17,78	18,13	19,50	19,09	20,26	—	19,67	95,48
August	18,73	17,32	19,92	19,37	20,46	—	19,91	96,65
Septembrie	22,53	19,49	23,88	19,47	20,96	—	20,21	98,10
Octombrie	23,70	25,83	27,57	19,55	21,33	—	20,44	99,22
Noiembrie	24,78	28,30	31,47	19,57	21,48	—	20,52	99,61
Decembrie	25,53	30,03	26,26	19,60	21,69	—	20,64	100,19

Sursa : *Anuarul statistic al R.P.R.* 1960.

După cum se vede din tabelul 46 producţia lunară de carne în anii 1958, 1959 şi 1960 prezintă pronunţate oscilaţii sezoniere. În consecinţă, pentru relevarea tendinţei de bază a producţiei de carne pe anii 1958—1960, s-a calculat seria mediilor mobile cu aplicarea formulei (XIII,2.1).

În figura 36 sînt reprezentate datele din tabelul 46. Se poate observa tendinţa centrală a producţiei de carne în R.P.R. în ultimii ani, marcînd o dezvoltare ascendentă.

Vom considera încă un exemplu (tabelul 47) pentru a ilustra calculul indicilor sezonalităţii.

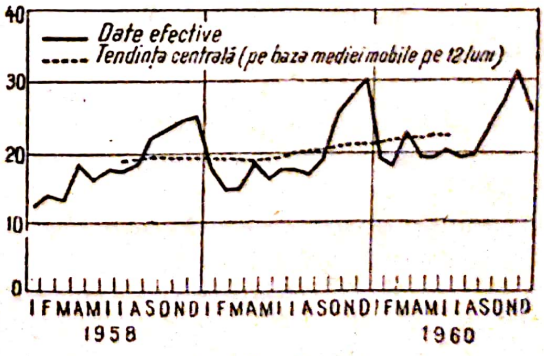


Fig. 36

După cum se vede din tabelul 47, transportul aerian de mărfuri are un caracter sezonier. Se observă, în fiecare an, o creştere a volumului de mărfuri transportate în lunile din mijlocul anului şi o scădere a acestuia în lunile de sfîrşit şi început ale fiecărui an.

Tabelul 47

Calculul indicilor sezonalității în transportul aerian de mărfuri al R.P.R. în perioada 1955—1959 (pe baza mediei aritmetice).

— mii tone-km —

Luna	1957	1958	1959	Medii lunare	Indici ai sezonalității
	1	2	3	4	5
Ianuarie	30	37	87	51,33	41,72
Februarie	45	51	112	69,33	56,36
Martie	54	50	141	81,67	66,39
Aprilie	60	73	118	83,67	68,01
Mai	80	72	144	98,67	80,21
Iunie	167	462	350	326,33	265,27
Iulie	111	269	252	210,91	175,29
August	92	147	158	152,43	107,57
Septembrie	86	232	159	159,00	129,15
Octombrie	52	87	180	106,33	86,43
Noiembrie	48	108	99	85,00	69,09
Decembrie	49	84	83	72,00	58,53
Total ...	874	1 672	1 883		
Media lunară	72,83	139,33	156,91	123,02	

Sursa : Buletin statistic trimestrial, *4* (1959).

Din punct de vedere statistic, se pune problema ca să fie măsurate oscilațiile sezoniere. Acest lucru se poate face prin două metode și anume : metoda mediei aritmetice ca în coloana 5 a tabelului 47 și metoda mediilor mobile ca în tabelul 48.

Indicii sezonalității din coloana 5 a tabelului 47 s-a obținut pe baza următoarelor operații :

— S-au calculat mediile pe fiecare lună (ianuarie, februarie etc.), considerîndu-se datele de pe rîndurile respective ale tabelului *), iar rezultatele sînt trecute în coloana 4 ;

— S-a calculat apoi media lunară generală pentru întreaga perioadă, în exemplul nostru $\bar{x} = 123{,}02$ mii tone-km.

— Și, în sfîrșit, s-au calculat indicii sezonalității, în procente, raportîndu-se mediile pe fiecare lună în parte la media lunară generală, iar rezultatele s-au înscris în coloana 5.

Media lunară generală în procente se echivalează cu 100 (se poate vedea ușor că media indicilor sezonalității este egală cu 100), ceea ce înseamnă că oscilațiile medii lunare încep acolo unde indicele sezonalității este mai mare sau mai mic de 100. O examinare sumară a datelor din coloana 5 a tabelului 47 ne arată că oscilațiile cele mai mari sînt în trimestrele II și III, atingînd un maxim în luna iunie, și cele mai mici sînt în trimestrele I și IV, cînd se înregistrează un minim de 41,72 mii tone-km.

Indicele sezonalității calculat prin metoda mediei aritmetice reflectă atît schimbările fenomenului de la an la an cît și oscilațiile sezoniere.

*) Datele din coloanele 1, 2, 3, formează o singură serie, formată din 36 termeni, care au fost aranjate pe trei coloane, pentru ușurarea calculelor.

Scanned with OKEN Scanner

Acest neajuns al metodei se atenuează dacă în serie nu există modificări simţitoare de la nivelul mediu general.

Pentru a putea urmări numai influenţa oscilaţiilor sezoniere la calculul indicilor sezonalităţii se foloseşte media mobilă.

Vom ilustra calculul indicilor sezonalităţii pe baza mediei mobile în tabelul 48.

Tabelul 48

Calculul indicilor sezonalităţii în transportul aerian de mărfuri al R.P.R. în perioada 1955—1959 (pe baza mediei mobile pe 12 luni)

Luna	Mediile mobile pe 12 luni ($\bar{y}_k$)			Indicatori ai abaterilor de sezon $\left(100\frac{y}{\bar{y}_k}\right)$			Mediile parţiale pe luni	Indici ai sezonalităţii (%)
	1957	1958	1959	1957	1958	1959		
	1	2	3	4	5	6	7	8
Ianuarie	—	105,17	155,87	—	35,18	55,81	45,50	46,91
Februarie	—	114,04	155,62	—	44,72	71,97	58,34	60,14
Martie	—	122,42	153,04	—	40,84	92,13	66,48	68,54
Aprilie	—	129,96	153,87	—	56,16	76,68	66,42	68,47
Mai	—	133,92	157,37	—	53,76	91,50	72,63	74,87
Iunie	—	137,87	156,96	—	335,10	222,98	279,04	287,66
Iulie	73,12	141,42	—	151,80	190,21	—	171,00	176,28
August	73,67	146,04	—	124,88	100,65	—	112,76	116,24
Septembrie	73,75	152,37	—	116,61	102,26	—	109,43	112,81
Octombrie	74,12	158,04	—	70,15	55,04	—	62,59	64,52
Noiembrie	74,33	162,92	—	64,57	66,29	—	65,43	67,45
Decembrie	86,29	161,25	—	56,77	52,09	—	54,43	56,11

$$\bar{x} = \frac{1164{,}05}{12} = 97$$

În coloanele 1—3 ale tabelului este calculată seria din medii mobile pe 12 luni, iar în coloanele 4—6, seria de indicatori de tipul celor arătaţi la (XIII, 2.2), în care se oglindeşte influenţa schimbării fenomenului în luna şi anul dat.

Mediile mobile s-au calculat pe baza formulei (XIII, 2.1), iar indicatorii de sezon prin raportarea termenilor originari ai seriei (vezi tabelul 47) la termenul corespunzător din seria ajustată cu ajutorul mediilor mobile. De pildă, indicele sezonal egal cu 151,80 (coloana 4, rîndul lunii iulie) s-a calculat raportîndu-se volumul mărfurilor transportate în luna iulie 1957 (111 mii tone-km) la media mobilă pe 12 luni egală cu 73,12 mii tone-km (această medie rezulta din :

$$\frac{\frac{30}{2} + 45 + 54 + 60 + 80 + 167 + \ldots + \frac{37}{2}}{12} = 73{,}12) .$$

În coloana 7 ne apar medii parţiale pe luni.

În sfîrșit, indicii sezonalității ne apar în coloana 8 prin raportarea fiecărei medii lunare, la media lunară generală, care în exemplul nostru este $\bar{x} = 97$.

Indicii din coloana 8 ne arată clar că valul sezonier are loc în lunile iunie-septembrie ale fiecărui an din perioada considerată; în aceste luni avem o depășire considerabilă peste 100.

Măsurarea oscilațiilor sezoniere prin indicii sezonalității este adeseori folosită în statistica social-economică.

O observație specială trebuie făcută în legătură cu indicatorii care oglindesc schimbările sezonale $\left(\frac{y}{\bar{y}_k}\right)$.

Într-o primă aproximare ei pot fi asimilați cu indicii sezonalității, însă, ei nu exprimă numai acțiunea factorilor sezonieri (condiții atmosferice bune sau rele pentru transportul aerian) ci și influența altor factori de organizare a transportului aerian ce se schimbă în decursul timpului. Dacă acești indicatori n-ar oglindi decît influența factorilor sezonieri, atunci nivelul lor într-o lună oarecare a anului — să zicem luna octombrie — ar rămîne aproximativ constant, de la an la an, ori, se poate observa chiar din exemplul nostru că indicatorii $100 \cdot \frac{y}{\bar{y}_k}$ pe luna aprilie oscilează de la nivelul cel mai mic de 53,76 (aprilie 1958) la nivelul de 91,50 (aprilie 1959).

Verificarea deplasărilor în valul sezonier. La elaborarea indicilor sezonalității statisticianul trebuie să țină seama de anumiți factori întîmplători care-și exercită influența chiar și în oscilațiile sezoniere cele mai stabile. Se pune, de aceea, problema de a face distincție între abaterile esențiale de cele neesențiale.

Să presupunem că indicatorul $100 \frac{y}{\bar{y}_k}$ ne-a dat pentru o lună oarecare, de exemplu ianuarie, valorile pe care le trecem în tabelul 49.

Tabelul 49

Anii	$100 \frac{y}{\bar{y}_k}$	*Rangul*
1946	104,75	2
1947	101,15	1
1948	117,21	4
1949	115,34	3
1950	126,45	5
1951	127,48	6
1952	136,64	8
1953	133,25	7
1954	147,23	9
1955	150,67	10
1956	152,33	12
1957	151,13	11

După cum se vede, indicatorul $100 \frac{y}{\bar{y}_k}$ ia valori cuprinse între 101,15 și 152,33. Aceste valori merg în general crescînd în perioada de 12 ani pe care am considerat-o; există însă

şi unele oscilaţii, astfel încît ne putem pune întrebarea dacă datele din coloana 2 a tabelului 49 reprezintă oscilaţii accidentale sau marchează într-adevăr o tendinţă de creştere.

Pentru a da răspunsul la această întrebare vom folosi coeficientul de corelaţie a rangurilor al lui Kendall (vezi p. 276).

Acest coeficient este egal cu

$$\rho = \frac{2(P-Q)}{n(n-1)}$$

iar dispersia lui este

$$D^2(\rho) = \frac{2(2n+5)}{9n(n-1)}.$$

Avem

$$P = 10+10+8+8+7+6+4+4+3+2+0 = 62,$$

$$Q = 1+0+1+0+0+0+1+0+0+0+1 = 4,$$

$$n = 12.$$

Prin urmare

$$\rho = \frac{2(62-4)}{12 \cdot 11} = 0{,}88,$$

$$D^2(\rho) = 0{,}0488.$$

Valoarea medie pătratică a lui ρ este $\sqrt{0{,}0488} = 0{,}22$.

Dacă valorile indicatorului s-ar fi grupat în mod accidental, valoarea lui ρ ar fi trebuit să fie apropiată de zero şi dată fiind valoarea dispersiei, cu o probabilitate foarte mare, să fie cuprinsă în intervalul $(-3D(\rho)),\ 3D(\rho)$, adică în intervalul $(-0{,}66,\ 0{,}66)$. Valoarea $\rho = 0{,}88$ este în afara intervalului şi deci putem conchide că există o tendinţă de creştere.

De remarcat e faptul că dacă am fi avut o creştere cu mult mai mare a valorilor indicatorului, de exemplu în loc de la 101,15 la 152,33, am fi avut de la 101,15 la 1 523,30, iar ordinea rangurilor din coloana 3 ar fi rămas aceeaşi, am fi ajuns la aceeaşi valoare a lui ρ. Coeficientul de corelaţie al rangurilor nu ne poate da deci o indicaţie asupra intensităţii creşterii, ci numai asupra faptului dacă există o creştere sau nu.

Din această cauză, coeficientul de corelaţie al rangurilor este cel mai indicat pentru a soluţiona problema pe care ne-am pus-o.

Metoda mediilor mobile are avantajul de a fi deosebit de elastică. Acest lucru se remarcă în special la studiul fenomenelor în care se manifestă schimbări brusce de intensitate în ritmul de creştere sau scădere, ca rezultat al schimbării condiţiilor lor de dezvoltare. Cînd au loc asemenea schimbări, media mobilă, cu însuşirea ei de a se adapta elastic noilor condiţii, poate constitui adeseori un mijloc mai eficient de determinare a tendinţei centrale decît folosirea unei funcţii matematice oarecare.

3. Metoda ajustării analitice. Alegerea curbei de tendinţă

Realitatea social-economică oferă statisticianului multe serii dinamice pentru care calculul tendinţei centrale este mai bine să se facă cu ajutorul curbelor matematice.

O metodă clasică pentru determinarea tendinței centrale este metoda *ajustării analitice.*

Cu ajutorul unei funcții matematice potrivite se stabilește direcția de bază a dezvoltării unui fenomen, ținîndu-se seama de complexitatea de factori care determină această direcție. Curba funcției, o linie dreaptă, o parabolă, hiperbolă etc., trebuie să fie, după cum s-a mai subliniat, *cea mai apropiată* de ansamblul de puncte extreme ale ordonatelor care conduc la historiograma seriei dinamice studiate. Acesta este principiul general al ajustării. Sarcina ajustării constă, deci, în a determina prin calcul o asemenea curbă.

Este evident că pentru aplicarea acestei metode trebuie să găsim, în primul rînd, ecuația curbei corespunzătoare care să înlocuiască datele reale ale seriei, satisfăcînd principiul de mai sus.

Ca și în toate problemele de folosire a calculului matematic în cercetarea statistică a fenomenelor social-economice, operația de ajustare a seriei trebuie să fie precedată de analiza teoretică a fenomenului respectiv, identificîndu-se factorii permanenți care-i determină existența și factorii întîmplători a căror acțiune determină oscilații neesențiale în dezvoltarea sa.

De asemenea, operația de ajustare trebuie să fie precedată și de o analiză *structurală* a șirului de date statistice. De exemplu, pentru cazul cînd se alege ecuația dreptei trebuie controlată existența unei anumite constanțe a creșterii (descreșterii) *absolute* de la un termen la altul al seriei; pentru curba exponențială, trebuie controlată existența unor creșteri (descreșteri) *relative*, aproximativ constante etc.

Pe baza acestor analize cercetătorul va decide tipul curbei de tendință pe care urmează s-o aleagă în operația de ajustare.

4. Ajustarea cu o linie dreaptă prin metoda celor mai mici pătrate

Ajustarea cu o linie dreaptă prin metoda celor mai mici pătrate este unul din procedeele relativ cele mai simple de ajustare. Acest procedeu — folosit pentru prima oară de marele matematician german K. F. Gauss în calculul compensării abaterilor — se bazează pe o variantă a principiului general al ajustării, formulat mai sus și anume : *suma pătratelor diferențelor dintre* y, *adică mărimile reale ale termenilor seriei, și* Y, *mărimile obținute prin ajustare* (*ordonatele corespunzătoare punctelor situate pe linia de tendință*) — *respectiv* $\Sigma(y - Y)^2$ — *trebuie să fie minimă.*

După cum se vede, ni se pune aceeași problemă ca și la studiul corelației liniare (vezi capitolul XI), cînd am considerat ecuația dreptei de forma

$$Y = b_1 x + a_1$$

Știm că metoda celor mai mici pătrate poate fi folosită la determinarea celor mai probabile valori ale lui a_1 și b_1 din ecuația dreptei. Prin analogie ajungem la concluzia că și în cazul cînd obiectul cercetării noastre statistice îl constituie o serie dinamică cu nivel liniar, putem să-i calculăm valorile probabile corespunzătoare ale constantelor din ecuația $Y = a + bt$ (în care t reprezintă timpul, iar Y variabila în timp cercetată).

Scanned with OKEN Scanner

Trebuie să atragem atenţia, însă, că deşi în activitatea teoretică şi practică de statistică ajustarea printr-o dreaptă prin metoda celor mai mici pătrate a găsit o largă aplicare şi în analiza variabilelor în timp, avînd o alură de dezvoltare liniară, premisele raţionale pe care se fundamentează această metodă n-au nimic comun cu seriile dinamice.

Într-adevăr, am precizat mai sus că indicatorii de nivel ai seriei dinamice *nu sînt valori independente* una faţă de alta şi, de aceea, abaterile de la curba funcţiei matematice ajustate sînt legate — în majoritatea cazurilor — de cauze ce nu au un caracter întîmplător.

Prin urmare, folosind metoda celor mai mici pătrate în studiul dinamicii fenomenelor ne ghidăm mai mult de consideraţii de ordin practic. Aplicarea metodei celor mai mici pătrate în anumite condiţii restrictive este, evident, admisibilă. Cu toate condiţiile restrictive în care o aplicăm, trebuie să nu pierdem din vedere că expresia matematică a nivelului dezvoltării nu reprezintă decît o funcţie *empirică*, practic utilă, şi nicidecum o funcţie matematică riguros exactă care ar exprima legile schimbării în timp ale fenomenelor fundamentate din punct de vedere teoretic. Aceasta înseamnă că determinarea prin mijloace matematice a nivelului de dezvoltare a fenomenelor social-economice se întemeiază pe ipoteza că există o anumită „lege" *) a dezvoltării fenomenului studiat; un asemenea punct de vedere este acceptabil dacă apreciem acea „lege" ca nefiind altceva decît manifestarea unei tendinţe constante, iar expresia ei matematică, ca mijloc, cel mai comod, de a o reprezenta sugestiv în perioada de timp studiată.

La ajustarea cu o dreaptă prin metoda celor mai mici pătrate ni se pune problema de a rezolva un sistem de două ecuaţii normale.

În cazul prelucrării seriilor dinamice, calculele pot fi întrucîtva simplificate. Dat fiind că t reprezintă de obicei un şir de numere succesive, originea poate fi fixată în punctul de mijloc. Dacă numărul termenilor seriei este impar, se înţelege că originea va fi fixată chiar în termenul de mijloc al seriei. Notînd abaterile de la origină cu t', vom avea evident $\Sigma t'$, egală cu 0, întrucît în acest caz, mărimile seriei care se distribuie pe axa y şi care reprezintă ordonata tendinţei — unele fiind cu semnul minus ($-$), altele cu semnul plus ($+$), se anulează reciproc. După cum se ştie, suma abaterilor de la media aritmetică este egală cu zero.

Astfel, sistemul de ecuaţii normale capătă următoarea formă:

$$\Sigma y = na, \qquad \text{(XIII, 4.1)}$$

$$\Sigma t'y = b\,\Sigma t'^2. \qquad \text{(XIII, 4.2)}$$

Din acest sistem rezultă că parametrii ecuaţiei tendinţei centrale sînt determinaţi de relaţiile:

$$a = \frac{\Sigma y}{n}\ ; \qquad b = \frac{\Sigma t'y}{\Sigma t'^2}\ . \qquad \text{(XIII, 4.3)}$$

*) Aici noţiunea de „lege" este dată în sensul ei strict matematic şi nu în sensul unei legi obiective social-economice care, după cum ştim, constituie obiectul de studiu al materialismului istoric şi al economiei politice marxiste.

Pe exemplul seriei dinamice oglindind dezvoltarea producției de sodă calcinată în R.P.R., în perioada 1951—1959, vom ilustra calculul tendinței centrale pe baza funcției liniare (vezi rezultatele calculului în tabelul 50).

Tabelul 50

Calculul tendinței centrale la producția de sodă calcinată a R.P.R. în perioada 1951—1959

Anii	Nr. de ordine a anilor (t)	Mărimi reale în mii tone (y)	Calculul tendinței liniare			
			Timp (t')	Produs ($t'y$)	t'^2	Tendința centrală ($y = a + bt$)
A	B	1	2	3	4	5
1951	1	49,5	−4	−198,0	16	37,90
1952	2	50,8	−3	−152,4	9	43,90
1953	3	55,1	−2	−110,2	4	49,90
1954	4	42,0	−1	− 42,0	1	55,80
1955	5	51,0	0	−502,6	0	61,80
1956	6	50,6	+1	+ 50,6	1	67,70
1957	7	67,4	+2	+134,8	4	73,70
1958	8	83,3	+3	+249,9	9	79,60
1959	9	106,1	+4	+424,4	16	85,50
Σ		555,8		+859,7	60	555,8

Sursa : După datele din *Anuarul statistic al R.P.R. 1959* și Buletinul statistic trimestrial, *4* (1959).

Tabelul 50 ne oferă datele necesare pentru calculul parametrilor ecuației liniare :

$$a = \frac{\Sigma y}{n} = \frac{555,8}{9} = 61,75,$$

$$b = \frac{\Sigma t'y}{\Sigma t'^2} = \frac{859,7 - 502,6}{60} = 5,95.$$

Ecuația liniară capătă forma

$$y = 61,75 + 5,95\,t.$$

Făcînd calculele corespunzătoare obținem valorile seriei ajustate

$$y_{1951} = 61,75 + 5,95\,(-4) = 37,90,$$

. . . . . .

$$y_{1959} = 61,75 + 5,95\,(+4) = 85,50.$$

Reprezentînd grafic (fig. 37) datele din coloana nouă a tabelului, observăm tendința centrală ilustrată printr-o dreaptă. Tot în acest grafic este ilustrată și tendința parabolică după calculele din tabelul 52 (p. 323) ca și tendința exponențială după calculele din tabelul 53 (p. 325).

Scanned with OKEN Scanner

Dacă numărul termenilor seriei este cu soţ, formulele (XIII, 4.3) pentru calculul parametrilor ecuaţiei rămîn aceleaşi, se schimbă doar anumite calcule de amănunt.

Să examinăm tabelul 51, unde avem calculată tendinţa centrală a vînzărilor de produse zaharoase în R.P.R., pe perioada 1951—1958. După cum se vede, seria noastră nu are un an median şi de aceea, pentru a simplifica calculele, vom plasa originea timpului în momentul care desparte anul 1954 de anul 1955, adică 1 ianuarie 1955. Mijlocul anului 1955 se va găsi la distanţa de şase luni de acest punct — în sens pozitiv — iar 1 ianuarie 1956, la distanţa de 12 luni etc.; în mod corespunzător, mijlocul anului 1954 se va găsi la distanţa de 6 luni de origine — în sens negativ — iar începutul anului 1953 la distanţa de 12 luni etc.

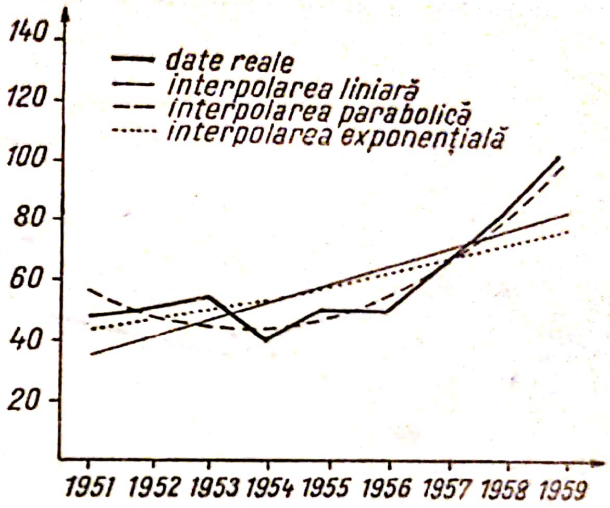


Fig. 37

Adoptînd ca unitate de măsură intervalul de timp egal cu şase luni, vom obţine pentru t' valorile din coloana 2 a tabelului 51.

Tabelul 51

Calculul tendinţei centrale la vînzările de zahăr prin comerţul socialist cu amănuntul în anii 1951—1958

Anii	Nr. de ordine al anilor (t)	Mărimi reale în mii tone (y)	Calculul tendinţei centrale			Tendinţa centrală ($y = a + bt$)
			Timp (t')	Produs ($t'y$)	t'^2	
A	B	1	2	3	4	5
1951	1	84,2	−7	−589,4	49	72,2
1952	2	76,1	−5	−380,5	25	80,4
1953	3	80,0	−3	−240,0	9	88,6
1954	4	85,0	−1	− 85,0	1	96,9
1955	5	111,0	+1	+111,0	1	105,1
1956	6	114,0	+3	+342,0	9	113,3
1957	7	128,5	+5	+642,5	25	121,5
1958	8	129,0	+7	+903,0	49	129,8
Σ	—	807,8	0	−1294,9 +1998,5	168	807,8

Sursa : Anuarul Statistic al R.P.R. 1959.

$$a = \frac{807,8}{8} = 100,97, \quad b = \frac{1\,998,5 - 1\,294,9}{168} = 4,19$$

Deci

$$y = 100,97 + 4,19\ t.$$

Calculele corespunzătoare ne dau seria de date din coloana 5, oglindind tendința vînzărilor de zahăr în R.P.R.

În figura 38 este reprezentată grafic această tendință. Trebuie să precizăm cu acest prilej că între parametrul b și coeficientul de regresie liniară (vezi capitolul XI) există o analogie formală. În cazul ajustării unei drepte prin metoda celor mai mici pătrate, parametrul b poate fi

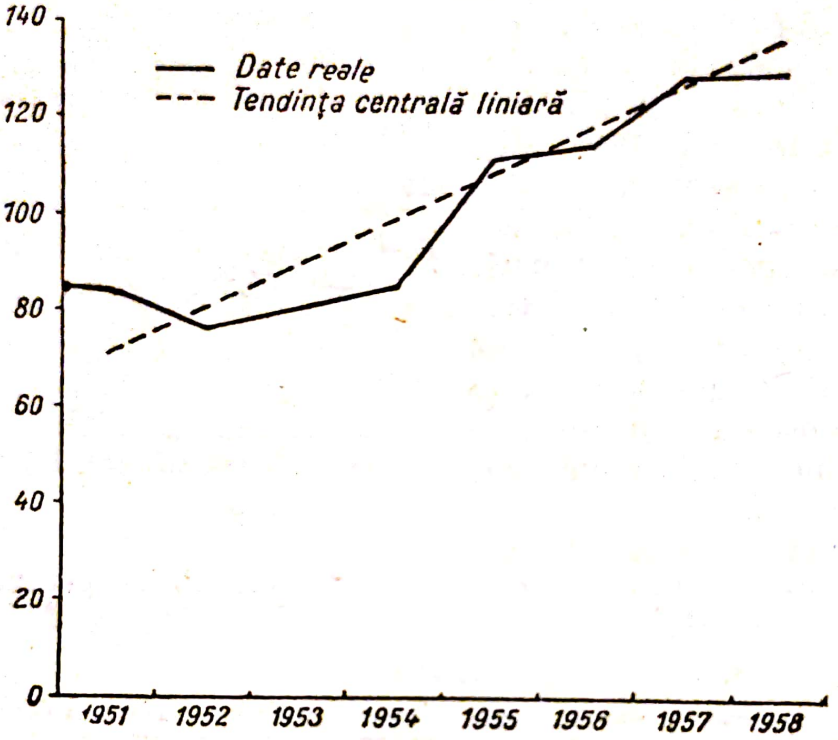


Fig. 38

privit ca un coeficient de regresie a lui y în t, fără să uităm, însă, că acest parametru nu are decît un suport formal spre deosebire de coeficientul de regresie care este o valoare aparținînd complexului ce definește o legătură statistică stabilă.

Coeficientului b i se mai spune panta liniei de tendință; valoarea lui reprezintă variația lui y corespunzătoare unei *unități de timp* datorită numai tendinței de dezvoltare a fenomenului.

Din al doilea exemplu considerat (vezi tabelul 51), s-a dedus că în perioada 1951—1958 volumul vînzărilor de zahăr prin comerțul socialist a crescut, în medie, cu $2.4{,}19 = 8{,}38$ mii tone anual (am văzut că valoarea lui b în acest exemplu se referă la un interval de 6 luni și de aceea pe o perioadă de un an ea se dublează).

Mai trebuie să atragem atenția că parametrul b poate fi interpretat ca fiind *viteza* de schimbare a seriei ajustate, întrucît viteza nu este altceva decît derivata de ordinul întîi a funcției de t.

În ultimul exemplu avem

$$y = 100{,}97 + 4{,}19\,t$$

sau

$$y' = 4{,}19. \qquad \text{(XIII, 4.4)}$$

Termenul constant al ecuaţiei liniare, care în exemplul nostru din tabelul 51 este egal cu 100,97, reprezintă nivelul ajustat al seriei în momentul de origine considerat (în speţă, în intervalul de timp de la 1 iulie 1954 la 1 iulie 1955), vînzările de zahăr în anul de origine se ridică la un nivel de 100,97 mii tone.

Întemeindu-ne pe aceste constatări, ecuaţia liniară ne dă posibilitatea să aflăm şi ritmul pentru orice moment al seriei ajustate.

Definind mai sus ritmul dezvoltării, am arătat că el reprezintă *mărimea sporului relativ pentru fiecare interval de timp al seriei.* De aceea, sporul relativ corespunzător unei unităţi de timp se exprimă prin formula

$$k = \frac{\Delta y}{y dt} \qquad \text{(XIII, 4.5)}$$

şi în cazul unei variaţii infinitezimale a variabilei t, el are următoarea expresie:

$$k = \frac{dy}{y dt} = \frac{y'}{y}. \qquad \text{(XIII, 4.6)}$$

Prin urmare, pentru a determina nivelul ritmului în fiecare moment al seriei ajustate, vom calcula raportul respectiv dintre derivata ecuaţiei dreptei şi termenul considerat din seria ajustată.

În exemplul din tabelul 51 nivelurile ritmului pentru anii succesivi 1951, 1952,... vor fi date de rapoartele:

$$\frac{4{,}19}{72{,}2} = 5{,}8\%; \qquad \frac{4{,}19}{80{,}4} = 5\%, \ldots \text{ etc.}$$

Asupra metodelor de calcul al ritmului *mediu* vom reveni mai jos.

5. Alte metode de ajustare

A. *Tendinţa parabolică.* Procedeele de ajustare printr-o dreaptă sînt cazuri particulare de ajustare prin curba parabolică.

Curba parabolică reprezintă expresia următoarei ecuaţii generale:

$$y = A_0 + A_1 t + A_2 t^2 + \ldots + A_{n-1} t^{n-1} + A_n t^n \qquad \text{(XIII, 5.1)}$$

După cum se vede, este vorba de o ecuaţie polinomială în t al cărei grad este determinat de cel mai mare exponent al puterii „variabilei"*) independente t. În particular, ecuaţia dreptei este tot o ecuaţie parabolică în care exponentul cel mai mare al puterii lui t este egal cu unu.

Ajustarea printr-o curbă parabolică de gradul 2 sau 3 etc. se face tot prin metoda celor mai mici pătrate, alcătuindu-se un sistem de $n+1$ ecuaţii normale pentru determinarea coeficienţilor:

$$A_0, A_1, \ldots, A_{n-1}, A_n.$$

*) Denumirea de „variabilă" independentă ce se atribuie lui t poartă aici un caracter formal, întrucît numărul de ordine a anilor nu poate fi considerat drept variabilă.

Dacă ne oprim, de exemplu, asupra unei ecuații parabolice de gradul doi, trebuie să construim un sistem de trei ecuații normale.

Folosind notația adoptată și la calculul tendinței liniare vom avea

$$y = a + bt + ct^2, \qquad \text{(XIII, 5.2)}$$

$$\sum y = na + b\sum t + c\sum t^2,$$

$$\sum ty = a\sum t + b\sum t^2 + c\sum t^3, \qquad \text{(XIII, 5.3)}$$

$$\sum t^2y = a\sum t^2 + b\sum t^3 + c\sum t^4.$$

Pentru simplificare, alegem de asemenea axa y în așa fel ca originea să cadă în mijlocul seriei, în care caz $\Sigma t = 0$ și $\Sigma t^3 = 0$, întrucît abaterile egal depărate de mijlocul seriei sînt egale și de semn contrar.

Considerînd această condiție, sistemul de ecuații (XIII, 5.3) capătă următoarea formă :

$$\left.\begin{aligned} \sum y &= n\,a + c\sum t^2, \\ \sum ty &= b\sum t^2, \\ \sum t^2y &= a\sum t^2 + c\sum t^4. \end{aligned}\right\} \qquad \text{(XIII, 5.4)}$$

Întrucît ecuația a doua conține numai o singură necunoscută b, ea poate fi determinată direct :

$$b = \frac{\Sigma ty}{\Sigma t^2}. \qquad \text{(XIII, 5.5)}$$

Sistemul format din prima și a treia ecuație ne dă posibilitatea să aflăm necunoscutele a și c, aplicînd regula lui Cramer, determinantul sistemului fiind diferit de zero.

Vom avea deci :

$$a = \frac{\begin{vmatrix} \Sigma y & \Sigma t^2 \\ \Sigma t^2y & \Sigma t^4 \end{vmatrix}}{\begin{vmatrix} n & \Sigma t^2 \\ \Sigma t^2 & \Sigma t^4 \end{vmatrix}} = \frac{\Sigma t^4 \cdot \Sigma y - \Sigma t^2 \cdot \Sigma t^2y}{n \cdot \Sigma t^4 - (\Sigma t^2)^2}, \qquad \text{(XIII, 5.6)}$$

$$c = \frac{\begin{vmatrix} n & \Sigma y \\ \Sigma t^2 & \Sigma t^2y \end{vmatrix}}{\begin{vmatrix} n & \Sigma t^2 \\ \Sigma t^2 & \Sigma t^4 \end{vmatrix}} = \frac{n\,\Sigma t^2y - \Sigma y\,\Sigma t^2}{n\,\Sigma t^4 - (\Sigma t^2)^2}. \qquad \text{(XIII, 5.7)}$$

Folosind din nou datele privind producția de sodă calcinată din R.P.R., vom întocmi un tabel de calcul pentru necunoscutele a, b, și c, ce ne sînt necesare (tabelul 52).

Scanned with OKEN Scanner

Tabelul 52

Calculul tendinţei parabolice la producţia de sodă calcinată a R.P.R. (1951—1958)

Anii	t	t^2	t^4	y	ty	t^2y	Tendinţa parabolică y
0	1	2	3	4	5	6	7
1951	−4	16	256	49,5	−198	792,0	55,4
1952	−3	9	81	50,8	−152,4	457,2	48,3
1953	−2	4	16	55,1	−110,2	220,4	44,9
1954	−1	1	1	42,0	− 42,0	42,0	45,3
1955	0	0	0	51,0	0	0	49,3
1956	1	1	1	50,6	50,6	50,6	57,1
1957	2	4	16	67,4	134,8	269,6	68,7
1958	3	9	81	83,3	249,9	749,7	83,9
1959	4	16	256	106,1	424,4	1 697,6	102,9
		60	708	555,8	357,1	4 279,1	555,8

$$a = \frac{708 \cdot 555{,}8 - 60\ 4\,279{,}1}{9 \cdot 708 - 60 \cdot 60} = \frac{393\,506 - 256\,746}{6\,372 - 3\,600} = \frac{136\,760}{2\,772} = 49{,}336,$$

$$b = \frac{357{,}1}{60} = 5{,}951,$$

$$c = \frac{9 \cdot 4\,279{,}1 - 555{,}8 \cdot 60}{9 \cdot 708 - 3\,600} = \frac{38\,511{,}9 - 33\,348}{6\,372 - 3\,600} = \frac{5\,164}{2\,772} = 1{,}863.$$

$$Y = 49{,}34 + 5{,}951\ t + 1{,}863\ t^2.$$

Valoarea constantelor a, b, şi c fiind determinată, calculul tendinţei parabolice devine simplu :

$$y_{1951} = 49{,}34 + (5{,}951 \cdot (-4)) + 1{,}863 \cdot 16 = 55{,}4,$$

. . . . . .

$$y_{1959} = 49{,}34 + (5{,}951 \cdot 4) + 1{,}863 \cdot 16 = 102{,}9.$$

În coloana 6 a tabelului 52 este calculată seria tendinţei centrale parabolice de gradul doi, iar în figura 36 este dată şi reprezentarea ei grafică (de comparat această tendinţă cu celelalte două reprezentate în grafic).

Rezolvarea sistemului de ecuaţii cu patru sau mai multe necunoscute necesită un mare volum de operaţii de calcul aşa încît ajustarea prin polinoame de grade superioare apare practic ca lipsită de interes. Pe de altă parte, o curbă cu mai mulţi parametri deşi este mai sensibilă la orice schimbare a datelor seriei, este mai puţin potrivită pentru a oglindi tendinţa centrală pe o perioadă mai mare.

B. *Ajustarea printr-o ecuaţie logaritmică.* Pentru a oglindi mai sugestiv schimbările relative sau raporturile de mărime în timp ale unor fenomene, în statistică se folosesc, după cum ştim, şi reprezentările grafice

pe scară logaritmică sau monologaritmică care s-au dovedit a fi foarte utile în practică.

Dacă, trecînd la scara logaritmică, datele seriei reale sînt dispuse aproximativ în linie dreaptă, vom putea folosi ecuaţia logaritmică pentru a releva tendinţa fenomenului. Formulele curbei liniare şi parabolice considerate pînă acum suferă transformări logaritmice corespunzătoare; în loc de y va apare $\lg y$. Aceasta înseamnă că în locul ecuaţiei dreptei vom avea

$$\lg y = a + bt,$$

iar în locul formulei generale a ecuaţiei polinomiale (parabolice) vom avea ecuaţia

$$\lg y = a + bt + ct^2 + dt^3 + \ldots$$

Curbele astfel obţinute pot fi reprezentate pe o reţea aritmetică obişnuită, sau monologaritmică (valorile şirului de numere naturale t şi logaritmii lui y), scara logaritmică fiind construită pe axa y.

Aplicînd logaritmii, o curbă complexă poate fi adusă la forma curbei liniare. Acest lucru se poate învedera considerînd, de exemplu, funcţia $y = x^2$. În sistemul coordonatelor rectangulare această funcţie, reprezentată grafic, ne dă o parabolă, pe cînd logaritmată, ea capătă forma:

$$\lg y = 2 \lg x$$

în care variabilele $\lg x$ şi $\lg y$ sînt combinate liniar.

C. *Ajustarea cu o curbă exponenţială prin metoda celor mai mici pătrate.* Un alt tip de curbă după care se calculează în mod obişnuit tendinţele centrale de dezvoltare ale anumitor fenomene este curba exponenţială.

Ecuaţia acestei curbe este

$$y_t = ab^t. \qquad \text{(XIII, 5.8)}$$

Aplicarea acestei ecuaţii în operaţia de ajustare poate da rezultate bune dacă dinamica seriei va indica o constanţă a creşterii *relative*.

Logaritmînd relaţia (XIII, 5.8), ea apare sub formă liniară în t:

$$\lg y = \lg a + t \lg b. \qquad \text{(XIII, 5.9)}$$

Vom considera din nou exemplul privind producţia de sodă calcinată a R.P.R. în perioada 1951—1958, calculînd de data aceasta tendinţa exponenţială (tabelul 52).

Tendinţa exponenţială o calculăm aplicînd la formula (XIII, 5.9) metoda celor mai mici pătrate.

Obţinem astfel:

$$\lg a = \frac{\Sigma \lg y}{n}, \quad \lg b = \frac{\Sigma t \lg y}{\Sigma t^2}.$$

Din calculele efectuate în tabelul 53, rezultă în cazul examinat

$$\lg a = \frac{15{,}9516108}{9} = 1{,}772402,$$

$$\lg b = \frac{2{,}2247035}{60} = 0{,}0370783.$$

Am obţinut astfel

$$\lg Y = 1{,}772402 + 0{,}0370783\, t.$$

Dînd lui t valorile din coloana 2, rezultă valorile respective ale logaritmilor valorilor ajustate, pe care le-am trecut în coloana 6. Cunoscînd

Tabelul 53

Calculul tendinţei exponenţiale la producţia de sodă calcinată în R.P.R., în perioada 1951—1959

Anii	y	t	t^2	$\lg y$	$t \lg y$	$\lg Y$	Y
A	1	2	3	4	5	6	7
1951	49,5	−4	16	1,6946052	−6,7784208	1,6240884	42,1
1952	50,8	−3	9	1,7058637	−5,1175911	1,6611668	45,8
1953	55,1	−2	4	1,7411516	−3,4823032	1,6982452	49,9
1954	42,0	−1	1	1,6232493	−1,6232493	1,7353236	54,4
1955	51,0	0	0	1,7075702	0	1,7724020	59,2
1956	50,6	1	1	1,7041505	1,7041505	1,8094804	64,5
1957	67,4	2	4	1,8286599	3,6573198	1,8465588	73,2
1958	83,3	3	9	1,9206450	5,7619360	1,8836372	76,5
1959	106,1	4	16	2,0257154	8,1028616	1,9207156	83,2
				15,9516108	2,2247035		

logaritmii, determinăm valorile finale, pe care le trecem în coloana 7. Cînd am aplicat ajustarea cu o curbă polinomială, am efectuat o verificare a calculelor bazîndu-ne pe egalitatea sumei valorilor reale cu suma valorilor ajustate.

De data aceasta avem în general

$$\Sigma Y \neq \Sigma y.$$

Verificarea calculelor o putem face însă egalînd media geometrică a valorilor reale cu media geometrică a valorilor ajustate.

Într-adevăr, în egalitatea

$$\lg Y = \lg a + t \lg b$$

dînd lui t valorile $0 \pm 1, \pm 2, \pm 3, \pm 4$ şi adunînd relaţiile obţinute, rezultă

$$\lg a = \frac{\Sigma \lg Y}{9}.$$

Însă am arătat că

$$\lg a = \frac{\Sigma \lg y}{9}.$$

Deci

$$\Sigma \lg Y = \Sigma \lg y$$

şi prin urmare

$$\sqrt[9]{\prod_{i=1}^{9} y_i} = \sqrt[9]{\prod_{i=1}^{9} Y_i}.$$

În cazul nostru avem

$$\sqrt[9]{49,5 \times 50,8 \times 55,1 \times 42,0 \times 51,0 \times 50,6 \times 67,4 \times 83,3 \times 106,1} = 59,2.$$

$$\sqrt[9]{42,1 \times 45,8 \times 49,9 \times 54,4 \times 59,2 \times 64,5 \times 73,2 \times 76,5 \times 83,2} = 59,2.$$

Prin urmare, ajustarea a fost bine făcută.

Calculînd antilogaritmii termenilor seriei din coloana 6, obţinem termenii corespunzători ai seriei din coloana 7, oglindind tendinţa centrală prin ajustarea după o curbă exponenţială. Această curbă este reprezentată grafic, alături de tendinţa liniară şi parabolică, în figura 37.

O întrebare firească se pune: care anume din cele trei variante de exprimare a tendinţei centrale (liniară, parabolică şi exponenţială) este mai corespunzătoare dezvoltării reale a fenomenului ilustrat — producţia de sodă calcinată în perioada 1951—1959.

Pentru a răspunde mai exact la această întrebare, vom verifica principiul care stă la baza aplicării metodei celor mai mici pătrate.

$$\Sigma (y - Y)^2 = \text{minim}.$$

Vom folosi în acest calcul schema din tabelul 54.

Tabelul 54

Anii	y	Y_1	Y_2	Y_3	$(y - Y_1)^2$	$(y - Y_2)^2$	$(y - Y_3)^2$
0	1	2	3	4	5	6	7
1951	49,5	37,9	55,4	42,1	134,56	134,81	54,76
1952	50,8	43,9	48,3	45,8	47,61	6,25	25,00
1953	55,1	49,9	44,5	49,9	27,04	104,04	27,04
1954	42,0	55,8	45,3	54,4	190,44	10,89	153,72
1955	51,0	61,8	49,3	59,2	116,44	2,89	67,24
1956	50,6	67,7	57,1	64,5	292,41	4,25	193,21
1957	67,4	73,7	68,7	73,2	39,69	1,69	33,64
1958	83,3	79,6	83,9	76,5	13,69	0,36	46,24
1959	106,1	85,5	102,9	83,2	424,36	10,24	524,41
					1 286,44	213,42	1 125,32

Se constată că suma pătratelor abaterilor (cea mai mică) este egală cu 213,42. Rezultă că tendinţa parabolică este cea mai apropiată de datele considerate.

La acelaşi rezultat ajungem şi dacă ne referim la valoarea absolută a abaterilor. Suma abaterilor absolute pentru ajustarea parabolică este 35,6, pentru ajustarea exponenţială 87,6 şi pentru ajustarea liniară 96. Prin urmare şi acest criteriu ne arată că cea mai bună ajustare a fost obţinută tot prin folosirea unei parabole.

Această concluzie este evident valabilă numai pentru seria particulară de date considerată. În general, nu se poate afirma că o ajustare duce

la un rezultat mai bun decît alta, decît după ce se face o verificare a mărimii abaterilor, aplicîndu-se diferite criterii, dintre care cele mai obişnuite constă în compararea sumei pătratelor abaterilor şi sumei valorilor absolute ale abaterilor, cum am făcut mai sus.

D. *Alte tipuri de curbe folosite la ajustare.* 1°. În practica statisticii şi planificării, problema calculului matematic al tendinţei unui fenomen social-economic capătă aspecte variate, după cum variate sînt şi nevoile reale de cunoaştere şi prevedere ştiinţifică a evoluţiei fenomenelor ce au loc în diferite ramuri ale producţiei materiale. Se poate ajusta, de exemplu evoluţia preţului de cost în funcţie de creşterea volumului producţiei, folosindu-se în acest scop ecuaţia hiperbolică.

Dependenţa preţului de cost de volumul producţiei poate fi exprimată prin următoarea relaţie:

$$y = a + \frac{b}{x}$$

în care:

a reprezintă volumul cheltuielilor de producţie variabile (proporţionale) pe unitate de produs şi

b volumul cheltuielilor convenţional-constante.

Calculul mărimilor a şi b nu este întotdeauna practic posibil, se ivesc o serie de dificultăţi legate de evidenţa lor exactă, ca şi de alte măsuri tehnico-organizatorice. Să ne referim, de exemplu, la cheltuielile legate de consumul de energie electrică la o întreprindere industrială. În lipsa unor contoare speciale, se înregistrează de obicei întregul consum. Or, o parte din cheltuielile ocazionate cu acest prilej se referă la cheltuieli constante cum ar fi iluminatul etc., iar o altă parte se referă la cheltuieli proporţionale cu volumul producţiei cum ar fi transportul cu ascensoare electrice a produselor finite, transportul intern de uzină electrificat etc.

Metoda celor mai mici pătrate ne oferă posibilitatea ca şi în acest caz să determinăm cu o precizie destul de mare, parametrii a şi b.

În tabelul 55 se dă un exemplu — cu date convenţionale — în care sînt ilustrate operaţiile de calcul ce ne sînt necesare.

Tabelul 55

Producţia de ciment a unei întreprinderi (în sute tone pe perioade de cîte 5 zile) şi calculul, cu ajutorul hiperbolei, al preţului de cost aferent volumului respectiv al producţiei

Producţia în sute tone (x)	Consumul de energie electrică pe o tonă de ciment în kWh (y)	$\frac{1}{x}$	$\left(\frac{1}{x}\right)^2$	$\left(\frac{1}{x}\right) \cdot y$
0	1	2	3	4
4	82	0,250	0,062500	20,50
5	74	0,200	0,040000	14,80
6	67	0,166	0,027556	11,12
6,5	72	0,153	0,023716	11,09
7	70	0,143	0,020449	1,001
Σ	365	0,913	0,174221	68,51

Pentru calculul parametrilor a şi b, avem nevoie de un sistem de două ecuaţii şi anume :

$$an + b\sum\frac{1}{x} = \Sigma y,$$

$$a\sum\frac{1}{x} + b\sum\left(\frac{1}{x}\right)^2 = \sum\left(\frac{1}{x}\right)y. \qquad \text{(XIII, 5.10)}$$

Aplicînd regula lui Cramer, aflăm pentru a şi b următoarele valori :

$$a = \frac{\begin{vmatrix} \Sigma y & \Sigma\frac{1}{x} \\ \Sigma\left(\frac{1}{x}\right)y & \Sigma\left(\frac{1}{x}\right)^2 \end{vmatrix}}{\begin{vmatrix} n & \Sigma\frac{1}{x} \\ \Sigma\frac{1}{x} & \Sigma\left(\frac{1}{x}\right)^2 \end{vmatrix}} = \frac{\Sigma y\cdot\Sigma\left(\frac{1}{x}\right)^2 - \Sigma\frac{1}{x}\cdot\Sigma\left(\frac{1}{x}\right)y}{n\Sigma\left(\frac{1}{x}\right)^2 - \left(\Sigma\frac{1}{x}\right)^2};$$

$$b = \frac{\begin{vmatrix} n & \Sigma y \\ \Sigma\frac{1}{x} & \Sigma\left(\frac{1}{x}\right)y \end{vmatrix}}{\begin{vmatrix} n & \Sigma\frac{1}{x} \\ \Sigma\frac{1}{x} & \Sigma\left(\frac{1}{x}\right)^2 \end{vmatrix}} = \frac{n\Sigma\left(\frac{1}{x}\right)y - \Sigma y\cdot\Sigma\frac{1}{x}}{n\Sigma\left(\frac{1}{x}\right)^2 - \left(\Sigma\frac{1}{x}\right)^2}.$$

Pe baza datelor din tabel avem :

$$a = \frac{63{,}590665 - 62{,}54963}{0{,}871105 - 0{,}833569} = \frac{1{,}041035}{0{,}037536} = 27{,}73430,$$

$$b = \frac{342{,}55 - 333{,}24}{0{,}871105 - 0{,}833569} = \frac{9{,}31}{0{,}037536} = 248{,}0285.$$

Introducînd aceste rezultate în ecuaţia funcţiei hiperbolice :

$$\overline{Y}_x = 27{,}73430 + \frac{248{,}0285}{x}.$$

vom putea determina consumul de energie electrică necesitat de creşterea volumului producţiei.

În tabelul 56 aceste cheltuieli sînt prezentate în ordinea creşterii producţiei.

Tabelul 56

x	4	5	6	6,5	7
a	27,734	27,734	27,734	27,734	27,734
$\frac{b}{x}$	62,007	49,854	41,173	38,196	35,468
$\overline{Y}$	89,741	77,588	68,907	65,930	63,202

Avem $\Sigma \overline{Y}_x = 365{,}368$, ceea ce arată exactitatea calculelor, diferenţa de 0,368 se datoreşte faptului că nu s-a lucrat cu un număr mai mare de zecimale.

2°. Curbele parabolice de tipul $y = at^b$, în general, nu sînt aplicabile seriilor dinamice ce caracterizează fenomenele social-economice întrucît timpul — variabila independentă — este considerată în aceste ecuaţii ca o progresie geometrică. Asemenea ecuaţii devin liniare dacă sînt logaritmate:

$$\lg y = \lg a + b \lg t.$$

În practică s-a observat că nivelurile seriilor economice se deosebesc uneori de progresiile geometrice printr-o constantă. Adăugînd sau scăzînd această constantă putem obţine o curbă pur exponenţială. În acest caz nivelul seriei apare sub forma

$$y = ab^t - k,$$

în care k este constanta cu care se abate seria reală de la progresia geometrică.

3°. O aplicaţie frecventă — îndeosebi în asigurări — a găsit curba lui Gompertz de forma:

$$y = ab^{c^t}. \qquad \text{(XIII, 5.11)}$$

Această curbă mai poate fi denumită „curba creşterii" întrucît oglindeşte creşterea cumulată a fenomenului pînă la un punct maximal oarecare. Această mişcare are loc în aşa fel încît la treptele superioare, creşterile absolute se micşorează, însă procesul creşterii continuă pînă la o limită anumită.

În scopul ajustării, ecuaţia curbei se transformă în forma ei logaritmică.

$$\lg y = \lg a + (\lg b)\, c^t. \qquad \text{(XIII, 5.12)}$$

Logaritmul lui a reprezintă valoarea maximă, „plafonul" către care tinde curba lui Gompertz; termenul al doilea din partea dreaptă a relaţiei (XIII, 5.12) reprezintă distanţa care separă nivelul maxim al seriei de nivelul ei din momentul dat (pe parcursul timpului această distanţă, bineînţeles, se micşorează); c va avea valoarea cuprinsă între 0 şi 1.

Scanned with OKEN Scanner

Determinarea curbei lui Gompertz se face cu oarecare aproximație dat fiind faptul că metoda celor mai mici pătrate în forma ei obișnuită nu este aplicabilă.

Determinarea constantelor din legea Gompertz-Makeham se poate face folosind oricare din metodele ajustării analitice. Formula exponențială a funcției de supraviețuire conducînd însă la calcule complicate, prin metodele clasice de ajustare, este nevoie să se aplice o ajustare specială.

Se pleacă de la :

$$l_x = ks^x g^{c^x}$$

și prin aplicarea logaritmilor, se obține

$$\lg l_x = \lg k + x \lg s + c^x \lg - g .$$

Șirul de vîrste date se împarte apoi în patru serii :

$$x,\ x+1,\ \ldots,\ x+t-1,$$
$$x+t,\ x+t+1,\ \ldots,\ x+2t-1,$$
$$x+2t,\ x+2t+1,\ \ldots,\ x+3t-1,$$
$$x+3t,\ x+3t+1,\ \ldots,\ x+4t-1.$$

Considerînd prima serie, avem :

$$\lg l_x = \lg k + x \lg s + c^x \lg g,$$
$$\lg l_{x+1} = \lg k + (x+1) \lg s + c^{x+1} \lg g,$$
$$\lg l_{x+2} = \lg k + (x+2) \lg s + c^{x+2} \lg g,$$
$$\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$
$$\lg l_{x+t-1} = \lg k + (x+t-1) \lg s + c^{x+t-1} \lg g.$$

Prin urmare, ținînd seama că se formează progresii aritmetice și geometrice, rezultă

$$u_x = t \lg k + \frac{t(2x+t-1) \lg s}{2} + \frac{c^x (c^t - 1)}{c-1} \lg g,$$

unde prin u_x notăm suma :

$$u_x = \lg l_x + \lg l_{x+1} + \ldots + \lg l_{x+t-1}.$$

A doua grupă se obține înlocuind pe x cu $x+t$. Așa că avem imediat

$$u_x = t \lg k + \frac{t}{2}(2x+t-1) \lg s + \frac{1}{c-1} c^x (c^t - 1) \lg g,$$
$$u_{x+t} = t \lg k + \frac{t}{2}(2x+3t-1) \lg s + \frac{1}{c-1} c^{x+t} (c^t - 1) \lg g,$$
$$u_{x+2t} = t \lg k + \frac{t}{2}(2x+5t-1) \lg s + \frac{1}{c-1} c^{x+2t} (c^t - 1) \lg g,$$
$$u_{x+3t} = t \lg k + \frac{t}{2}(2x+7t-1) \lg s + \frac{1}{c-1} c^{x+3t} (c^t - 1) \lg g.$$

Scanned with OKEN Scanner

Din teoria diferenţelor se ştie că :

$$\Delta u_x = u_{x+1} - u_x.$$

Aplicînd această formulă, vom scădea pe u_x din u_{x+t} etc. şi vom obţine

$$\Delta u_x = t^2 \lg s + \frac{c^x(c^t-1)^2}{c-1} \lg g,$$

$$\Delta u_{x+t} = t^2 \lg s + \frac{c^{x+t}(c^t-1)^2}{c-1} \lg g,$$

$$\Delta u_{x+2t} = t^2 \lg s + \frac{c^{x+2t}(c^t-1)^2}{c-1} \lg g.$$

Acestea sînt diferenţele de ordinul întîi. Trecînd acum la diferenţele de ordinul al doilea :

$$\Delta^2 u_x = \frac{c^x(c^t-1)^3}{c-1} \lg g,$$

$$\Delta^2 u_{x+t} = \frac{c^{x+t}(c^t-1)^3}{c-1} \lg g.$$

Se obţine apoi prin împărţire :

$$\frac{\Delta^2 u_{x+t}}{\Delta^2 u_x} = c^t.$$

Deducem imediat pe c.

Cunoscînd pe c deducem pe g în mod succesiv din relaţiile scrise.

Determinăm astfel valorile g, s, k, c, pornind de la valorile brute ale tabelelor de mortalitate şi legea lui Gompertz-Makeham este în întregime cunoscută.

4°. O curbă ce s-a folosit adeseori în lucrările de statistică demografică este aşa-numita „curbă logistică" care are multe asemănări cu curba lui Gompertz. Ea reprezintă o progresie geometrică modificată în care creşterea îşi micşorează intensitatea pe măsura apropierii de o limită anumită.

Una din formele acestei curbe ce poate fi folosită comod la determinarea tendinţei se caracterizează prin următoarea ecuaţie :

$$\frac{1}{y} = a + bc^t. \qquad \text{(XIII, 5.13)}$$

Observăm că această ecuaţie este ecuaţia modificată a curbei exponenţiale, cu singura deosebire că variabila dependentă este aici $\frac{1}{y}$ şi nu y.

Metoda de ajustare se întemeiază pe principiile ce ne sînt cunoscute, introducîndu-se însă mărimile inverse ale lui y.

6. Cîteva considerații generale asupra ajustării. Formule de interpolare

Am arătat pînă acum diferite tipuri de curbe ce pot fi folosite la ajustarea seriilor în timp, atrăgînd atenția încă de la începutul acestui capitol asupra importanței pe care o are alegerea curbei de tendință într-un caz concret sau altul. Trebuie să mai precizăm acum că în soluționarea corectă a acestei probleme cuvîntul hotărîtor îl are experiența cercetătorului, întrucît un criteriu obiectiv unic pentru orice serie concretă ce o supunem cercetării nu poate fi dat.

Și într-adevăr, este dificilă dacă nu imposibilă formularea unor principii generale după care să ne orientăm în rezolvarea problemei.

Există, totuși — în afară de analiza economică multilaterală și de structură — unele elemente auxiliare care ne pot ajuta la alegerea celei mai potrivite curbe de tendință.

În primul rînd, este indicată întocmirea mai multor reprezentări grafice cît mai corespunzătoare caracterului datelor statistice cuprinse în serie (pe o simplă rețea aritmetică sau rețea mono-logaritmică, rețea logaritmică dublă etc.). Asemenea grafice ne vor putea sugera cu suficientă aproximație, curba teoretică ce trebuie folosită la ajustare.

În al doilea rînd, o curbă potrivită se poate alege examinîndu-se în prealabil raportul reciproc în care se află variabilele t și y. De exemplu, dacă variația lui t ia forma progresiei aritmetice, iar valorile corespunzătoare ale lui y formează o progresie geometrică, relația dintre variabile este ilustrată de o ecuație exponențială

$$y = ab^t,$$

sau, dacă ambele variabile formează serii de progresie geometrică, relațiile dintre ele vor fi de tip parabolic sau hiperbolic

$$(y = at^b) \text{ etc.}$$

În al treilea rînd, practic pot fi încercate mai multe curbe iar alegerea curbei de ajustare să se facă comparîndu-se în prealabil rezultatele obținute. Dacă ecuațiile curbelor comparate conțin un număr egal de parametri, compararea pătratelor abaterilor de la nivelurile seriei inițiale ne poate servi drept criteriu suficient pentru alegerea curbei.

Suma pătratelor abaterilor poate fi calculată după următoarea formulă:

$$\sum (d^2) = \sum (y^2) - a \sum (y) - b \sum (ty) - c \sum (t^2 y) \ldots$$

Este de menționat că prin el însuși, gradul de apropiere a datelor reale de curba funcției folosită la ajustare nu constituie încă un criteriu sigur în alegerea curbei de tendință. Putem avea o curbă ce se apropie foarte mult de datele reale ale seriei și care să nu reprezinte totuși tendința centrală a dezvoltării fenomenului. Nu trebuie să uităm că noțiunea de tendință centrală este identică cu noțiunea unei curbe line de la care există abateri însă care determină tendința unei mișcări de *lungă durată*; aceasta

înseamnă că o astfel de curbă, satisfăcînd, geometric, conceptul de tendinţă centrală, trebuie să fie cît mai simplă şi totodată cît mai corespunzătoare datelor reale.

De asemenea, trebuie avute în vedere şi limitele de folosire a curbei de tendinţă. În general, elementele pe care le avem în vedere aici şi care ne ajută să alegem curba de tendinţă sînt recomandabile numai în cazul cînd curba căutată ne va servi la interpolarea datelor în limitele seriei reale şi nicidecum pentru a efectua eventualele extrapolări. O curbă bine aleasă pentru interpolări, nu poate fi întotdeauna potrivită şi pentru extrapolări, *decît în ipoteza că sîntem siguri că şi în viitor condiţiile materiale de dezvoltare a fenomenului studiat vor rămîne neschimbate.*

În al patrulea rînd, ne putem găsi în faţa unei serii căreia să nu-i corespundă nici un tip de curbă din cele folosite în mod obişnuit întrucît alura dezvoltării fenomenului studiat are forme cu totul neregulate, ca urmare a schimbărilor ce intervin în dezvoltarea fenomenului studiat. În acest caz este util ca cercetătorul să separe datele seriei iniţiale în cîteva perioade ce prezintă particularităţi asemănătoare de dezvoltare şi să le ajusteze pe fiecare în parte, pe baza unor curbe corespunzătoare.

Este evident că în asemenea cazuri se pierde perspectiva de ansamblu asupra seriei şi de aceea, folosirea procedeului de divizare a seriei iniţiale trebuie făcută cu mult discernămînt, ea contrazicînd cerinţele categoriei statistice de tendinţă centrală.

În general, în formula pe care am ales-o pentru ajustare va intra un număr oarecare de parametri ale căror valori se determină în funcţie de datele experimentale. Adeseori aceasta ne obligă să rezolvăm anumite sisteme de ecuaţii care dau naştere la calcule complicate. De aceea, în practică, se impune fixarea numărului acestor parametri ceea ce va avea o mare importanţă pentru prezentarea fenomenului pe cale analitică.

Formule de interpolare

1°. Formula de interpolare a lui Lagrange *). Să determinăm un polinom de gradul n;

$$P(x) = a_0 x^n + a_1 x^{n+1} + \ldots + a_{n-1} x + a_n, \qquad \text{(XIII, 6.1)}$$

care pentru $x = x_i$, ia valorile $Y_i = P(x_i)$, unde $i = 0, 1, 2, \ldots, n$.

Polinomul va fi determinat dacă vom cunoaşte coeficienţii $a_0, a_1, \ldots, a_n$. Pentru aceasta, scriem cele $n + 1$ ecuaţii liniare care rezultă din datele problemei

$$\sum_{i=0}^{n} a_i x_j^{n-i} = y_j, \qquad (j = 0, 1, 2, \ldots, n).$$

De aici găsim soluţiile

$$a_i = \sum_{0}^{n} A^{(i)} \cdot y_j, \qquad (j = 0, 1, 2, \ldots, n)$$

*) J. L. Lagrange (1726—1813), matematician francez, a făcut descoperiri importante în matematică, mecanică şi astronomie.

de unde rezultă că coeficienții a_i se exprimă liniar în funcție de valorile date $y_0, y_1, \ldots, y_n$. Vom putea deci scrie relația:

$$P(x) = y_0 P_0(x) + y_1 P_1(x) + \ldots + y_n P_n(x) = \sum_0^n {}_k \; P_k(x) y_k, \qquad \text{(XIII, 6.2)}$$

unde $P_k(x)$ sînt polinoame de gradul n pe care urmează să le determinăm. Deoarece $P(x_0) = y_0$, relația precedentă se scrie:

$$y_0 P_0(x_0) + y_1 P_1(x_0) + \ldots + y_n P_n(x_0) = y_0$$

de unde rezultă valorile

$$P_0(x_0) = 1$$
$$P_i(x_0) = 0$$

pentru $i = 1, 2, \ldots, n$.

În mod analog, obținem grupele de relații:

$$P_1(x_1) = 1, \qquad P_i(x_1) = 0, \qquad (i = 0, 2, \ldots, n)$$
$$P_2(x_2) = 1, \qquad P_i(x_2) = 0, \qquad (i = 0, 1, 3, \ldots, n)$$
$$\ldots\ldots\ldots\ldots\ldots$$
$$P_n(x_n) = 1, \qquad P_i(x_n) = 0, \qquad (i = 0, 1, \ldots, n-1)$$

Avem, deci, pentru polinomul $P_0(x)$, condițiile

$$P_0(x_0) = 1,$$
$$P_0(x_i) = 0, \qquad (i = 1,2, \ldots, n).$$

Deoarece polinomul $P_0(x)$ este de gradul n și se anulează pentru $x = x_1$, $x = x_2$, $\ldots$, $\ldots, x = x_n$, putem scrie, notînd cu C_0 o constantă

$$P_0(x) = C_0 \prod^{(0)} (x - x_i),$$

unde $\prod^{(0)} (x - x_i)$ reprezintă produsul

$$(x - x_1)(x - x_2) \ldots (x - x_n),$$

din care lipsește factorul $x - x_0$, adică

$$\prod^{(0)} (x - x_i) = (x - x_1)(x - x_2) \ldots (x - x_n).$$

Cum $P_0(x_0) = 1$, avem:

$$P_0(x_0) = C_0 \prod^{(0)} (x_0 - x_i) = 1.$$

De aici deducem valoarea coeficientului C_0:

$$C_0 = \frac{1}{\prod^{(0)} (x_0 - x_i)}.$$

Prin urmare,

$$P_0(x) = \frac{\prod^{(0)} (x - x_i)}{\prod^{(0)} (x_0 - x_i)} = \frac{(x - x_1)(x - x_2) \ldots (x - x_n)}{(x_0 - x_1)(x_0 - x_2) \ldots (x_0 - x_n)}.$$

În general, vom avea:

$$P_k(x) = \frac{(x-x_0)(x-x_1)\dots(x-x_{k-1})(x-x_{k+1})\dots(x-x_n)}{(x_k-x_0)(x_k-x_1)\dots(x_k-x_{k-1})(x_k-x_{k+1})\dots(x_k-x_n)}$$

sau cu notaţiile prescurtate de mai sus:

$$P_k(x) = \frac{\overset{(k)}{\prod}(x-x_i)}{\overset{(k)}{\prod}(x_k-x_i)}, \qquad (k \neq i).$$

Formula de interpolare a lui Lagrange este, deci, înlocuind aceste valori în (XIII, 6.2):

$$\boxed{P(x) = \sum_0^n{}_k \frac{\overset{(k)}{\prod}(x-x_i)}{\overset{(k)}{\prod}(x_k-x_i)} y_k, (k \neq i)} \qquad \text{(XIII, 6.3)}$$

Această formulă se poate pune şi sub o altă formă şi anume, notînd

$$F(x) = (x-x_0)(x-x_1)\dots(x-x_n)$$

obţinem

$$P(x) = \sum_0^n{}_n \left[\frac{F(x)}{F'(x_k)}\right] \frac{y_k}{x-x_k}. \qquad \text{(XIII, 6.4)}$$

2°. Un caz particular al formulei de interpolare a lui Lagrange, se obţine atunci cînd cunoaştem, pentru două valori ale variabilei independente, valorile corespunzătoare ale funcţiei.

Luînd, deci, în formula (XIII, 6.3) pe $n = 1$, obţinem:

$$P(x) = \frac{x-x_1}{x_0-x_1} y_0 + \frac{x-x_0}{x_1-x_0} y_1,$$

care se poate scrie sub o formă mai uşor de calculat:

$$P(x) = y_0 + \frac{y_0-y_1}{x_0-x_1}(x-x_0). \qquad \text{(XIII, 6.5)}$$

Din punct de vedere geometric a aplica formula (XIII, 6.5) înseamnă a înlocui curba necunoscută care trece prin punctele de coordonate $P(x_0, y_0)$, $Q(x_1, y_1)$ prin coarda PQ.

Punctul intermediar de abscisă x $(x_0 < x < x_1)$ va avea ordonata $P(x)$ dată de formula (XIII, 6.5).

3°. *Formula de interpolare a lui Newton* *). Să considerăm $n+1$ numere în progresie aritmetică

$$x_0, x_0 + r, \dots, x_0 + nr \qquad (\alpha)$$

*) Isac Newton (1642—1727), unul din marii oameni de ştiinţă ai lumii. El a formulat legile mecanicii clasice. A descoperit legea gravitaţiei universale şi legile de descompunere a luminii. În acelaşi timp cu Leibnîtz, Newton a pus bazele calculului diferenţial şi integral, primul plecînd de la consideraţii geometrice, iar al doilea de la consideraţii mecanice. În anul 1687 a publicat opera lui monumentală: *Philosophiae naturalis principia mathematica* (publicată şi în limba romînă: Isac Newton, *Principiile matematice ale filozofiei naturale* — Introducere şi adnotări de prof. Victor Marian, Ed. Acad. R.P.R., Bucureşti, 1956).

și valorile corespunzătoare ale funcției necunoscute y:

$$y_0, y_1, \ldots, y_n. \qquad (\beta)$$

Din teoria diferențelor se știe că

$$y_k = y_0 + \frac{k}{1!}\Delta y_0 + \frac{k(k-1)}{2!}\Delta^2 y_0 + \ldots + \Delta^k y_0.$$

Vom scrie atunci:

$$y_k = y_0 + \frac{k}{1!}\Delta y_0 + \frac{k(k-1)}{2!}\Delta^2 y_0 + \ldots + \frac{k(k-1)\ldots(k-k+1)}{k!}\Delta^k y_0.$$

Dacă notăm

$$k = \frac{x - x_0}{r}$$

rezultă

iar formula de mai sus devine, notînd cu $P(x)$ valoarea corespunzătoare a lui y_n:

$$P(x) = y_0 + \frac{x - x_0}{r}\cdot\frac{\Delta y_0}{1!} + \frac{x - x_0}{r}\left(\frac{x - x_0}{r} - 1\right)\frac{\Delta^2 y_0}{2!} + \ldots$$
$$\ldots + \left(\frac{x - x_0}{r}\right)\left(\frac{x - x_0}{r} - 1\right)\ldots\left(\frac{x - x_0}{r} - n + 1\right)\frac{\Delta^n y_0}{n!}. \qquad \text{(XIII, 6.6)}$$

Se vede că y este un polinom în x de gradul n și ia valorile (β) atunci cînd x ia valorile (α). Așadar formula (XIII, 6.6) reprezintă un polinom de gradul n care, pentru $n + 1$ valori în progresie aritmetică date variabilei independente x, ia $n + 1$ valori date. El aproximează o funcție necunoscută $f(x)$ care are aceleași proprietăți ca și polinomul $P(x)$.

Formula (XIII, 6.6) se numește *formula de interpolare a lui Newton*. De reținut este faptul că spre deosebire de formula lui Lagrange pentru care valorile variabilei independente sînt numere oarecare, în formula lui Newton, valorile x_i sînt numai în progresie aritmetică. Se vede deci, că formula lui Lagrange este de aplicație mai generală.

4°. *Formula lui Newton generalizată.* Am văzut că formula lui Newton se aplică în cazul în care variabila independentă funcției ia valori în progresie aritmetică. Deci, dacă notăm prin $x_0, x_1, \ldots, x_n$ valorile variabilei independente, avem diferențele:

$$\Delta x_0 = \Delta x_1 = \ldots = \Delta x_{n-1} = r.$$

În cazul în care valorile Δx_i, ($i = 0, 1, 2, \ldots, n$), sînt diferite între ele, adică intervalele sînt neegale, vom aplica o formulă asemănătoare cu formula lui Newton, pe care o numim formula lui Newton generalizată și pe care o vom stabili mai departe. Mai înainte, însă, vom introduce o noțiune nouă, și anume: noțiunea de „diferențe divizate".

Fie $f(x)$ o funcție oarecare și $x_0, x_1, \ldots, x_n$ valorile diferite între ele ale variabilei independente x. Numim diferențe divizate de ordinul întîi ale funcției $f(x)$ rapoartele:

$$f(x_1, x_0) = \frac{f(x_1) - f(x_0)}{x_1 - x_0}, \qquad f(x_2, x_1) = \frac{f(x_2) - f(x_1)}{x_2 - x_1},$$
$$f(x_3, x_2) = \frac{f(x_3) - f(x_2)}{x_3 - x_2},$$

diferențe divizate de ordinul al doilea:

$$f(x_2, x_1, x_0) = \frac{f(x_2, x_1) - f(x_1, x_0)}{x_2 - x_0},$$
$$f(x_3, x_2, x_1) = \frac{f(x_3, x_2) - f(x_2, x_1)}{x_3 - x_1}, \ldots,$$

diferenţe divizate de ordinul al treilea:

$$f(x_3, x_2, x_1, x_0) = \frac{f(x_3, x_2, x_1) - f(x_2, x_1, x_0)}{x_3 - x_0}.$$

Diferenţele divizate de orice ordin ale unei funcţii sînt simetrice în raport cu valorile variabilei independente.

Demonstraţia se face prin inducţie, observînd că $f(x_1, x_0)$ se poate scrie sub forma

$$f(x_1, x_0) = \frac{f(x_0)}{x_0 - x_1} + \frac{f(x_1)}{x_1 - x_0} = f(x_0, x_1).$$

Avem de asemenea

$$f(x_2, x_1, x_0) = \frac{f(x_2, x_1) - f(x_1, x_0)}{x_2 - x_0} = \frac{\dfrac{f(x_2) - f(x_1)}{x_2 - x_1} - \dfrac{f(x_1) - f(x_0)}{x_1 - x_0}}{x_2 - x_0} =$$

$$= \frac{1}{x_2 - x_0} \cdot \left[\frac{f(x_1)}{x_1 - x_2} + \frac{(fx_2)}{x_2 - x_1} - \frac{f(x_0)}{x_0 - x_1} - \frac{f(x_1)}{x_1 - x_0} \right]$$

sau încă, efectuînd calculele,

$$f(x_2, x_1, x_0) = \frac{f(x_0)}{(x_0 - x_1)(x_0 - x_2)} + \frac{f(x_1)}{(x_1 - x_2)(x_1 - x_0)} + \frac{f(x_2)}{(x_2 - x_0)(x_2 - x_1)}$$

deci putem scrie egalităţile

$$f(x_2, x_1, x_0) = f(x_1, x_2, x_0) = \dots$$

În general, admitem relaţia

$$f(x_n, x_{n-1}, \dots, x_1, x_0) = \sum_{k=0}^{n} \frac{f(x_k)}{(x_k - x_0) \dots (x_k - x_{k-1})(x_k - x_{k+1}) \dots (x_k - x_n)} \qquad (a)$$

şi să demonstrăm că pentru $(n + 2)$ numere x_k subzistă relaţia

$$f(x_{n+1}, x_n, \dots, x_1, x_0) = \sum_{k=0}^{n+1} \frac{f(x_k)}{(x_k - x_0) \dots (x_k - x_{k-1})(x_k - x_{k+1}) \dots (x_k - x_{n+1})}. \qquad (b)$$

Într-adevăr, avem prin definiţie

$$f(x_{n+1}, x_n, \dots, x_1, x_0) = \frac{f(x_{n+1}, x_n, \dots, x_1) - f(x_n, \dots, x_1, x_0)}{x_{n+1} - x_0}. \qquad (c)$$

Pe baza ipotezei (a), avem însă relaţiile

$$f(x_{n+1}, x_n, \dots, x_1) = \sum_{k=1}^{n+1} \frac{f(x_k)}{(x_k - x_1) \dots (x_k - x_{k-1})(x_k - x_{k+1}) \dots (x_k - x_{n+1})},$$

$$f(x_n, \dots, x_1, x_0) = \sum_{k=0}^{n} \frac{f(x_k)}{(x_k - x_0) \dots (x_k - x_{k-1})(x_k - x_{k+1}) \dots (x_k - x_n)},$$

de unde rezultă, înlocuind în relaţia (c),

$$f(x_{n+1}, x_n, \dots, x_1, x_0) =$$

$$= \frac{1}{x_{n+1} - x_0} \sum_{k=1}^{n} \frac{f(x_k)}{(x_k - x_1) \dots (x_k - x_n)} \left[\frac{1}{x_1 - x_{n+1}} - \frac{1}{x_k - x_0} \right].$$

Cum avem însă

$$\frac{1}{x_k - x_{n+1}} - \frac{1}{x_k - x_0} = \frac{x_{n+1} - x_0}{(x_k - x_0)(x_k - x_{n+1})}$$

efectuînd calculele obţinem relaţia (b), deci relaţia (a) este generală. Putem deci scrie:

$$f(x_n, x_{n-1}, \dots, x_1, x_0) = f(x_{n-1}, x_n, \dots, x_1, x_0) = \dots,$$

adică toate cele $P_{n+1} = (n + 1)!$ valori ale funcţiei f sînt egale. Deci, $f(x_0, x_1, \dots, x_n)$ este o funcţie simetrică de valorile $x_0, x_1, \dots, x_n$.

Avem tabelul diferenţelor divizate (tabelul 57):

Tabelul 57

x	$f(x)$	I	II	III	...
x_0	$f(x_0)$				
x_1	$f(x_1)$	$f(x_0, x_1)$			
x_2	$f(x_2)$	$f(x_1, x_2)$	$f(x_0, x_1, x_2)$		
x_3	$f(x_3)$	$f(x_2, x_3)$	$f(x_1, x_2, x_3)$	$f(x_0, x_1, x_2, x_3)$	
x_4	$f(x_4)$	$f(x_3, x_4)$	$f(x_2, x_3, x_4)$	$f(x_1, x_2, x_3, x_4)$	
⋮	⋮	⋮	⋮	⋮	

Să calculăm acum valoarea funcţiei $f(x)$ pentru $x = t$. Diferenţa de ordinul n fiind constantă, putem scrie:

$$f(x_0, x_1, .., x_n) = f(t, x_0, x_1, \ldots, x_{n-1}) \tag{d}$$

dar, avem prin definiţie:

$$f(t,x_0,x_1,\ldots, x_{n-1}) = \frac{f(t, x_0, \ldots, x_{n-2}) - f(x_0, x_1, \ldots, x_{n-1})}{t - x_{n-1}},$$

de unde

$$f(t, x_0, x_1, \ldots, x_{n-1})\,(t - x_{n-1}) + f(x_0, x_1, \ldots, x_{n-1}) = f(t, x_0, \ldots, x_{n-2})$$

sau, ţinînd seama de relaţia (d),

$$f(t, x_0, \ldots, x_{n-2}) = f(x_0, x_1, \ldots, x_{n-1}) + (t - x_{n-1})\, f(x_0, x_1, \ldots, x_n). \tag{e}$$

Putem scrie succesiv:

$$f(t, x_0, \ldots, x_{n-2}) = \frac{f(t, x_0, \ldots, x_{n-3}) - f(x_0, \ldots, x_{n-2})}{t - x_{n-2}}$$

$$f(t, x_0, \ldots, x_{n-3}) = f(x_0, \ldots, x_{n-2}) + (t - x_{n-2})\, f(t, x_0, \ldots, x_{n-2}).$$

Înlocuind în această relaţie pe $f(t, x_0, \ldots, x_{n-2})$ din (e), obţinem:

$$f(t, x_0, \ldots, x_{n-3}) = f(x_0, \ldots, x_{n-2}) + (t - x_{n-2})\, f(x_0, \ldots, x_{n-1}) +$$
$$+ (t - x_{n-1})(t - x_{n-2})\, f(x_0, x_1, \ldots, x_n).$$

Din aproape în aproape se ajunge la formula lui Newton generalizată:

$$f(t) = f(x_0) + (t - x_0)\, f(x_0, x_1) + (t - x_0)(t - x_1)\, f(x_0, x_1, x_2) + \ldots$$
$$\ldots + (t - x_0)(t - x_1) \ldots (t - x_{n-1})\, f(x_0, x_1, \ldots, x_n). \tag{f}$$

EXEMPLUL 42. O funcţie ia valorile

$$300;\ 784;\ 2\,904;\ 4\,732;\ 19\,404,$$

respectiv pentru $x = 10; 14; 22; 26; 42$. Să se calculeze valoarea funcţiei pentru $x = 12{,}84$.

Vom forma tabelul diferenţelor divizate (tabelul 58).

Tabelul 58

x	$f(x)$	I	II	III
10	300			
			24	
14	784	121	32	1
22	2 904	265	46	1
26	4 732	457		
42	19 404	917		

Aplicînd formula (f), în care facem $t = 12{,}84$, obţinem

$$f(12{,}84) = 594{,}75.$$

Scanned with OKEN Scanner

Capitolul XIV

Metode de calcul al ritmului mediu de dezvoltare

1. Generalităţi

Pentru caracterizarea ritmului mediu de dezvoltare a fenomenelor social-economice, statistica foloseşte metode şi procedee diferenţiate, în funcţie de particularităţile fenomenelor cercetate şi de sarcinile analizei statistice.

După cum se ştie, pentru conducerea planificată a economiei naţionale, urmărirea ritmului de dezvoltare a diferitelor ei ramuri, prezintă o importanţă deosebită. În planurile anuale şi de perspectivă se stabilesc nivelurile pe care trebuie să le atingă pînă la sfîrşitul perioadei planificate producţia globală, producţia în expresie naturală, volumul investiţiilor etc., ca şi ritmurile corespunzătoare acestora. În această privinţă, un exemplu strălucit ni-l oferă documentele istorice ale Congresului al III-lea al P.M.R., privind planul de dezvoltare a economiei naţionale pe anii 1960—1965 ca şi schiţa programului economic de perspectivă pe 15 ani (perioada 1960—1975). Organizarea muncii şi caracterizarea economică şi statistică a rezultatelor obţinute se fac pe baza posibilităţilor reale ale întreprinderilor care trebuie să asigure desfăşurarea producţiei în ritm corespunzător pentru ca la sfîrşitul perioadei sarcinile să fie realizate şi depăşite.

O lege generală a construcţiei socialismului şi comunismului, confirmată de experienţa tuturor ţărilor lagărului socialist, este ritmul rapid de dezvoltare a economiei şi culturii. Acţiunea ei se întemeiază pe proprietatea socială asupra mijloacelor de producţie, pe dezvoltarea armonioasă, planică, a tuturor ramurilor producţiei materiale, pe creşterea cu precădere a producţiei mijloacelor de producţie, pe relaţiile socialiste de producţie. Ele au influenţat atitudinea nouă faţă de muncă a producătorilor nemij-

lociți, gradul lor de cointeresare în ridicarea productivității muncii, au stimulat elanul creator al oamenilor muncii. Ritmul rapid de dezvoltare este deci caracteristic orînduirii socialiste; el indică superioritatea netăgăduită a socialismului față de capitalism.

Este evident, deci, că cercetarea statistică a ritmului de dezvoltare, determinarea corectă a categoriei statistice a *ritmului mediu* de dezvoltare a fenomenelor și proceselor social-economice, constituie una din sarcinile importante ale statisticii.

În munca statistică, ritmul mediu se utilizează în primul rînd, ca element de calcul pentru preliminări și elaborări de schițe-program de perspectivă, iar în al doilea rînd, ca mijloc de control pentru sesizarea mersului executării planului etc.

În capitolul precedent am arătat că ritmul reprezintă mărimea sporului *relativ* pentru fiecare interval de timp al seriei și că el poate fi exprimat prin formulele (XIII, 4.5) și (XIII, 4.6), în cazul unei ajustări după o dreaptă.

Pe de altă parte, am mai arătat că sporul relativ poate fi exprimat prin două categorii de indicatori: *coeficienți ai ritmului de creștere* și indicatori de *ritm al sporului*. Relația dintre acești indicatori este de forma arătată la (XIII, 4.5).

Cele două categorii de indicatori corespund, în general, conceptului de *ritm* care exprimă „desfășurarea gradată, treptată a unei acțiuni, evoluția mai rapidă sau mai lentă a unei activități condiționată de anumiți factori determinanți" *). Într-adevăr, desfășurarea „gradată, treptată" a fenomenelor rezultă cu claritate dacă se compară un nivel al seriei dinamice cu un altul, luat ca bază de comparație (nivelul precedent, cel inițial sau oricare altul). Comparația, evident, poate fi exprimată matematic fie printr-un raport direct între nivelurile considerate, în care caz obținem *ritmul de creștere* (descreștere), fie printr-o diferență dintre nivelurile considerate, raportată apoi la nivelul luat ca bază de comparație, în care caz obținem *ritmul sporului*.

În general, se consideră că un fenomen se dezvoltă în ritm uniform, cînd coeficienții individuali în formă de lanț sau sporurile relative între valorile succesive ale seriei *exprimă o anumită constanță*. O creștere uniformă nu admite o tendință categoric crescîndă (descrescîndă), ea nu exclude însă și unele abateri întîmplătoare, fie chiar mai mari **).

După cum vom vedea, toate metodele folosite în practică pentru calculul ritmului mediu se întemeiază pe această ipoteză.

Așadar, se pune problema de a determina, în modul cel mai potrivit pentru scopuri practice, ritmul mediu de dezvoltare a fenomenelor social-economice.

Ne vom opri, în primul rînd, asupra procedeelor de calcul elaborate pînă acum, la determinarea ritmului mediu de creștere (descreștere).

---

*) *Dicționarul limbii romîne*. Ed. Acad. R.P.R., București, 1958, p. 720.

**) Vezi în acest sens articolul *A mertani átlag statisztikai alkalmazásai* de P. Köves, în revista „Statisztikai Szemle", *4—5* (1957).

2. Metoda mediei geometrice

În practica statisticii socialiste, ritmul mediu se calculează prin media geometrică a coeficienților ritmului de creștere.

Am văzut în capitolul X că media geometrică se bazează pe relația care există între produsul termenilor unui șir de valori și termenul final al seriei. Aplicînd-o în cazul valorilor reprezentînd coeficienții de creștere pe care-i notăm cu k_i, putem scrie:

$$y_n = y_0 \cdot k_1 \cdot k_2 \ldots k_n .$$

Relația este satisfăcută dacă

$$y_0 \cdot k_1 \cdot k_2 \ldots k_n = y_0 \bar{k}^n .$$

Simplificînd cu y_0 ambele părți ale relației și extrăgînd rădăcina de ordinul n din prima parte a acesteia, obținem valoarea medie a coeficienților de creștere — ritmul mediu de creștere:

$$\bar{k} = \sqrt[n]{k_1 \cdot k_2 \ldots k_n} = \sqrt[n]{\prod_{i=1}^{n} k_i}, \quad (i = 1,2,\ldots, n). \quad \text{(XIV, 2.1)}$$

Folosind logaritmii, vom obține ușor nivelul ritmului mediu de creștere

$$\lg \bar{k} = \frac{\Sigma \lg k_i}{n} . \quad \text{(XIV, 2.2)}$$

Cunoscînd că produsul coeficienților în formă de lanț se reduce de fapt la coeficientul în formă de bază al ultimei perioade oglindite în serie, valoarea medie a ritmului de creștere capătă forma

$$\bar{k} = \sqrt[n]{\frac{y_n}{y_0}} \quad \text{(XIV, 2.3)}$$

sau în formă logaritmică

$$\lg \bar{k} = \frac{\lg y_n - \lg y_0}{n} . \quad \text{(XIV, 2.4)}$$

Împotriva folosirii mediei geometrice la determinarea ritmului mediu de creștere s-au adus mai multe obiecții care, în esență, se rezumă la faptul că această medie este sensibilă numai la nivelurile inițial și ultim al seriei și ignorează toți termenii intermediari. În consecință, în cazul oscilațiilor brusce ale indicatorilor economici — determinate în special de crizele economice, cum este cazul în economia capitalistă — calculul mediei geometrice pentru a determina ritmul mediu de dezvoltare a fenomenului considerat ar fi lipsit de orice sens *).

Cu anumite restricții, unii autori admit calculul ritmului mediu cu ajutorul mediei geometrice, dar numai în cazurile particulare ale desfă-

*) Vezi în acest sens I. Pisarev, *Nekotorie voprosi teorii statistiki*. Voprosi Ekonomiki, 7 (1948).

şurării fenomenului şi anume cînd coeficienţii ritmului prezintă, în general, o anumită constanţă, fie o creştere neîntreruptă, fie o scădere continuă)*. Se mai preconizează că media geometrică este aplicabilă numai pentru calcule de orientare şi numai în acele cazuri cînd ritmul mişcării se schimbă foarte puţin de la un an la altul.

În sfîrşit, unii autori condiţionează folosirea mediei geometrice de cunoaşterea condiţiilor reale ale dezvoltării fenomenului, cunoscut fiind faptul că ritmul mediu calculat prin media geometrică apare ridicat cînd ritmurile scad vertiginos de la an la an şi dimpotrivă, micşorat, cînd seria prezintă ritmuri din ce în ce mai înalte **).

3. Metoda mediei parabolice

În anul 1949, statisticianul sovietic A. Karasev ***) şi matematicianul romîn A. Froda ****) propun în mod independent ca ritmul mediu de creştere să fie calculat cu ajutorul *mediei parabolice*. Această medie ţine seama de suma valorilor termenilor seriei şi de valoarea termenului iniţial.

Dacă notăm cu θ coeficientul ritmului mediu de creştere, cu a_0 nivelul iniţial al seriei dinamice şi cu n numărul termenilor seriei, se ajunge la următoarea relaţie:

$$a_0 + a_0\theta + a_0\theta^2 + a_0\theta^3 + \ldots + a_0\theta^n = \Sigma a_i \qquad \text{(XIV, 3.1)}$$

din care, dînd pe a_0 factor comun, obţinem

$$a_0(1 + \theta + \theta^2 + \ldots + \theta^n) = \Sigma a_i .$$

Şi mai departe, împărţind ambele părţi ale egalităţii cu a_0, obţinem

$$1 + \theta + \theta^2 + \ldots + \theta^n = \frac{\Sigma a_i}{a_0} . \qquad \text{(XIV, 3.2)}$$

Întrucît prima parte a egalităţii (XIV, 3.2) reprezintă suma termenilor unei progresii geometrice, relaţia (XIV, 3.2) mai poate fi scrisă şi sub forma *****):

$$\text{(I)} \qquad \frac{\theta^{n+1} - 1}{\theta - 1} = \frac{\Sigma a_i}{a_0} . \qquad \text{(XIV, 3.3)}$$

Calculul ritmului mediu cu ajutorul acestei formule prezintă oarecare dificultăţi deoarece valoarea lui θ este stabilită prin calcule aproximative.

---

*) Vezi *Teoria Statisticii*, Ed. de stat pentru literatură economică şi juridică, Bucureşti, 1955.

**) În legătură cu discuţiile ce se poartă în problema folosirii mediei geometrice la calculul ritmului mediu ca şi în problema folosirii altor procedee de calcul, vezi articolul lui I. Marinescu, *Consideraţii teoretice cu privire la metodele de calcul al ritmului mediu de dezvoltare*. Statistica, 7 (1960).

***) A. Karasev, *K voprosu ob iscislenii spednegodovih tempov rosta naturinovo hozeaistva*. Vestnik Statistiki, 2 (1949).

****) A. Froda, *Numere-indice cumulatoare*. Analele Academiei R.P.R., III (1949).

*****) Vom folosi aici o notaţie suplimentară a formulelor cu cifre romane pentru a scoate în evidenţă şase formule mai importante.

Examinînd avantajele aplicării acestei metode în comparaţie cu cea a mediei geometrice, statisticianul chinez Fin-Di-Şi o aplică la un exemplu al dinamicii investiţiilor pe 5 ani *). El deduce următoarea serie de indici de bază :

$$1{,}05\ ;\ 1{,}39\ ;\ 1{,}44\ ;\ 1{,}71\ ;\ 2{,}64$$

pentru care dacă se calculează ritmul mediu cu ajutorul mediei geometrice — egală cu 1,2143 — dinamica investiţiilor este denaturată. Se ajunge la un volum cumulat al investiţiilor cu 12,92% mai mare decît cel real. O asemenea diferenţă nu este admisibilă.

Media parabolică pe care o notează cu q reprezintă ritm mediu care dacă este pus în locul ritmurilor individuale de mai sus nu denaturează volumul total al investiţiilor, fiind satisfăcută relaţia :

$$\bar{q} + \bar{q}^2 + \bar{q}^3 + \bar{q}^4 + \bar{q}^5 = 1{,}05 + 1{,}39 + 1{,}44 + 1{,}71 + 2{,}64$$

sau

$$\bar{q} + \bar{q}^2 + \bar{q}^3 + \bar{q}^4 + \bar{q}^5 - 8{,}23 = 0.$$

Din această ecuaţie rezultă că $\bar{q} = 1{,}171$. Practic, pentru calculul ritmului mediu parabolic se pot folosi tabele speciale care au fost elaborate în R.P. Chineză.

Aceste tabele au fost construite pentru calculul ritmului mediu de la 100,1% pînă la 150,0%, pe 2, 3, 4 sau 5 ani. Evident, ele pot fi construite şi pe un număr mai mare de ani.

Dăm în tabelul 59 un extras din aceste tabele.

Tabelul 59

Tabel pentru calculul ritmului mediu $\bar{q}$

Ritmul mediu căutat ($\bar{q}$)	Suma indicilor de bază efectivi			
	pe doi ani	pe trei ani	pe patru ani	pe cinci ani
1,1600	2,6056	4,0665	5,8771	7,9774
1,1610	2,5089	4,0738	5,8902	7,9999
1,1620	2,5122	4,0811	5,9042	8,0226
1,1630	2,5156	4,0887	5,9182	8,1459
1,1640	2,5189	4,0964	5,9317	8,0685
1,1650	2,5222	4,1033	5,9453	8,0912
1,1660	2,5256	4,1109	5,9594	8,1148
1,1670	2,5289	4,1182	5,9729	8,1373
1,1680	2,5322	4,1256	5,9867	8,1605
1,1690	2,5356	4,1382	6,0008	8,1840
1,1700	2,5389	4,1405	6,0114	8,2069
1,1710	2,5422	4,1479	6,0282	8,2300
1,1720	2,5456	4,1555	6,0423	8,2536
1,1730	2,5489	4,1628	6,0559	8,2765
1,1740	2,5523	4,1704	6,0700	8,3001
1,1750	2,5556	4,1778	6,0889	8,3236

*) Fin-Di-Şi, *K voprosu o metodologii iscislenia srednevo tempa rosta*, Vestnik Statistiki, 2(1958).

După acest tabel s-a calculat şi $\bar{q} = 1{,}171$ din exemplul luat de Fin--Di-Şi.

Un exemplu concludent dă şi Al. Froda *).

Fie dată statistic, prin şirurile S sau S' producţia în tone a unui produs anumit într-o perioadă cincinală; în ambele cazuri producţia din primul şi ultimul an fiind respectiv aceeaşi, adică

S :	$p_0 = 10\,000$		S' :	$p'_0 = 10\,000$
	$p_1 = 11\,000$			$p'_1 = 14\,000$
	$p_2 = 13\,200$			$p'_2 = 18\,200$
	$p_3 = 17\,160$			$p'_3 = 21\,840$
	$p_4 = 24\,024$			$p'_4 = 24\,024$

de unde producţia totală este respectiv 75 384 şi 88 064 :

$$P_0 = \sum_k^{0,4} p_k = 75\,384\,; \qquad P'_0 = \sum_k^{0,4} p'_k = 88\,064.$$

Calculînd indicii elementari, găsim respectiv

$$i_1 = 1{,}1\,; \quad i_2 = 1{,}2\,; \quad i_3 = 1{,}3\,; \quad i_4 = 1{,}4\,;$$
$$i'_1 = 1{,}4\,; \quad i'_2 = 1{,}3\,; \quad i'_3 = 1{,}2\,; \quad i'_4 = 1{,}1.$$

Indicii aritmetic şi geometric sînt pentru S, ca şi pentru S' :

$$i_a = \frac{1}{4}\,(1{,}1 + 1{,}2 + 1{,}3 + 1{,}4) = 1{,}25\,;$$

$$i_g = \sqrt[4]{1{,}1 \times 1{,}2 \times 1{,}3 \times 1{,}4} = 1{,}245.$$

Indicii rezultînd din aplicarea metodei parabolice sînt :

$$i = 1{,}206\,; \qquad i' = 1{,}285,$$

după cum se obţine cu aproximaţie pe baza unui calcul destul de uşor.

Metoda mediei parabolice a fost criticată şi ea **). S-a obiectat, în primul rînd, că ritmul mediu calculat după această metodă este orientat nu numai după suma tuturor nivelurilor seriei dinamice ci şi după nivelul primului termen, acordîndu-i-se acestuia preferinţă fără nici o argumentare. Datorită acestei metodologii nivelul ritmului mediu este direct proporţional cu suma valorilor seriei şi invers proporţional cu valoarea primului termen. Nu este greu de observat că dacă între valoarea primului termen şi celo următoare ale seriei există diferenţe mari, ritmul mediu apare mărit sau diminuat în mod corespunzător. Această metodă ar da eventual rezultate satisfăcătoare dacă fenomenele studiate au o dezvoltare aproximativ constantă. În al doilea rînd, s-a considerat că întrucît media para-

*) Al. Froda, *Numere indice cumulatoare*. Analele Academiei R.P.R., **III** (1949).

**) Vezi L. Kazineţ, *Ob iscislenii svodnih pokazatilei dinamiki sotialino-ekonomiceskii iavleniia*. Vestnik Statistiki, *5* (1956) şi P. Petrov, *Po voprosa za tempovete na riasta*. Statistika, *2* (1960).

bolică nu depinde decît de suma termenilor, nivelul ei nu este influențat dacă se schimbă ordinea valorilor seriei; de asemenea, ritmul mediu poate fi calculat prin media parabolică numai pentru seriile dinamice de intervale în care totalizarea valorilor seriei are un sens real *). În sfîrșit, s-a mai obiectat că în cazul cînd ritmul mediu s-ar calcula prin media parabolică, s-ar ajunge la situații cînd corelația de sistem dintre anumiți indicatori s-ar „destrăma", deoarece ritmul mediu nu poate fi calculat decît pentru unii indicatori din cadrul sistemului (L. Kazineț dă exemplul corelației care trebuie să existe între productivitatea muncii, numărul muncitorilor și volumul producției globale), iar pentru alți indicatori calculul lui nu ar avea sens.

O observație interesantă, demnă de reținut, a făcut statisticianul bulgar P. Petrov.

El consideră, de pildă, că dacă modelul matematic al dezvoltării fenomenului studiat reprezintă, în ultima formă, o funcție pătratică de tipul:

$$y = a + bt + ct^2$$

a cărei curbă este o parabolă de gradul doi, ea se caracterizează prin aceea că atît prima derivată a acesteia, cît și raportul acestei derivate față de funcția respectivă sînt mărimi variabile. De aici rezultă că atunci cînd schimbările respective au ca model o parabolă de gradul doi, calculul ritmului nu-și are sensul, deoarece *nu există acel ritm constant care să se apropie de ritmul mediu*. Aceeași situație — arată autorul — se întîlnește și în cazul cînd modelul matematic obținut este o parabolă de un grad mai mare.

4. Metoda lui Froda pentru calculul ritmului mediu de creștere

Determinarea ritmului mediu de creștere rezultă din ecuația (I), a cărei rezolvare, după cum am arătat, nu este simplă.

Prof. Al. Froda **) a arătat modul practic de rezolvare a acestei ecuații, pe baza unor considerații interesante nu numai din punctul de vedere matematic, ci și economic.

Vom reda, în cele ce urmează, fără modificări importante, analiza făcută de Al. Froda.

Să notăm în ecuația (I):

$$a_{on} = \frac{\Sigma a_i}{n+1};$$

a_{0n} este, prin urmare, producția medie în anii considerați, presupunînd că termenii seriei a_i reprezintă volumul unei producții oarecare.

*) Statisticianul maghiar P. Köves, referindu-se la acest argument al lui L. Kazineț, îl consideră ca neîntemeiat întrucît, pe de o parte, aplicabilitatea mediei parabolice nu poate fi condiționată de împrejurarea dacă analizăm o serie de intervale sau de momente, iar pe de altă parte, unele serii de momente pot fi interpretate ca serii de intervale. De pildă, seriile oglindind evoluția numărului mediu scriptic al muncitorilor sau productivitatea muncii pot fi concepute și ca serii dinamice de om-luni sau, respectiv, ca serii asupra producției raportate la unitatea de timp și ca atare, datele respective ar putea fi însumate.

**) *Loc. cit.*, p. 342.

Revenim la ecuația (I) pe care o scriem sub forma

$$(II) \qquad \frac{(1+\rho)^{n+1}-1}{(n+1)\rho} = \frac{a_{0n}}{a_0} = r,$$

unde raportul r arată în ce măsură a crescut (descrescut) producția anuală mijlocie față de cea inițială, iar $\theta = 1 + \rho$.

Se poate demonstra că pentru $r > 1$, $\rho > 0$, iar pentru $r < 1$, $\rho < 0$.

Într-adevăr, notînd $x = 1 + \rho$, ecuația (II) devine:

$$f(x) = x^{n+1} - (n+1)\,rx + [(n+1)\,r - 1] = 0. \qquad (XIV, 4.1)$$

Se vede imediat că ecuația admite rădăcina $x = 1$, care nu este însă o soluție acceptabilă deoarece primul membru al ecuației (II), nu are sens pentru $\rho = 0$. Dacă ρ tinde către zero în (II) prin valori pozitive sau negative, rezultă $r = 1$ și reciproc, astfel că $\rho = 0$ implică $r = 1$ și invers.

Avem întotdeauna

$$r > \frac{1}{n+1}$$

căci aceasta revine la $(n+1)\,a_{0n} > a_0$.

De aici rezultă că ecuația algebrică prezintă două variații, în sensul teoremei lui Descartes, și că are, deci, două rădăcini pozitive sau nici una. Deoarece una din rădăcinile pozitive este $x = 1$, urmează că cealaltă rădăcină pozitivă $\bar{x}$ dă adevărata soluție.

Vom aplica acum teorema lui Rolle.

Derivata a lui $f(x)$ este

$$f'(x) = (n+1)\,(x^n - r).$$

Există, după cum se vede, o singură rădăcină pozitivă $x' = \sqrt[n]{r}$ a derivatei.

Se poate ușor vedea că valorile negative ale lui ρ nu sînt acceptabile dacă $|\rho| \geqslant 1$. Într-adevăr, dacă $\rho = -1$ din relația

$$a_1 = a_0\,(1 + \rho),$$

rezultă $a_1 = 0$, iar dacă $\rho < -1$ din aceeași relație rezultă $a_1 < 0$. Este însă evident că nu avem producții negative sau egale cu zero. Rezultă, deci, că dacă $\rho < 0$ trebuie neapărat ca $|\rho| < 1$.

Dacă $r > 1$, avem $\sqrt[n]{r} > 1$, deci conform teoremei lui Rolle, rădăcina $\bar{x} > \sqrt[n]{r}$, așa încît

$$\rho > \sqrt[n]{r} - 1 > 0 \qquad (XIV, 4.2)$$

Dacă $r < 1$, avem $\sqrt[n]{r} < 1$, deci

$$-\rho > 1 - \sqrt[n]{r} > 0. \qquad (XIV, 4.3)$$

Proprietatea enunțată este deci adevărată.

Scanned with OKEN Scanner

Rezolvarea ecuației (I) se poate face prin metodele generale de aproximare ale algebrei și prin procedee specifice matematicii financiare. Într-adevăr, ecuația (I) se întîlnește în matematica financiară, unde $a_{\rho n}$ se consideră a fi o sumă acumulată prin depuneri anuale a_0 posticipate și se caută procentul anual posticipat ρ de fructificare a depunerilor *).

Metodele matematicilor financiare nu sînt însă aplicabile, deoarece ele se referă la valori pozitive ale lui ρ și în general la valorile mici ale lui ρ, deoarece ρ reprezintă procentul de fructificare al unei sume de bani. În cazul nostru însă ρ poate lua și valori negative, atunci cînd producția descrește de la un an la altul. De asemenea, ρ poate lua și valori mai mari decît 1, ceea ce în matematicile financiare nu se poate întîmpla.

Se impune deci rezolvarea ecuației (II) fără folosirea tabelelor financiare.

Scriem din nou ecuația (II) sub forma următoare:

$$\frac{x^n}{(n+1)r} + \frac{(n+1)r-1}{(n+1)rx} = 1.$$

Se vede că dacă $r > \frac{1}{n+1}$, rezultă

$$\frac{x^n}{(n+1)r} \geqslant 0, \qquad \frac{(n+1)r-1}{(n+1)rx} \geqslant 0.$$

Putem deci nota, ținînd seamă de relația precedentă,

$$\frac{x^n}{(n+1)r} = \cos^2\alpha, \qquad \frac{(n+1)r-1}{(n+1)rx} = \sin^2\alpha.$$

Eliminînd pe x între aceste două egalități, rezultă ecuația trigonometrică

$$\sin^{2n}\alpha \cos^2\alpha = \frac{[(n+1)r-1]^n}{[(n+1)r]^{n+1}}.$$

Aplicînd logaritmii rezultă:

$$n \lg \sin\alpha + \lg\cos\alpha = \frac{1}{2}[n \lg(nr+r-1) - (n+1)\lg(nr+r)].$$

Unghiul α se calculează acum folosind tabelele de logaritmi. Bineînțeles, nu găsim o valoare exactă a lui α ci o valoare aproximativă, cu atît mai exactă, cu cît tabelele de logaritmi au un număr mai mare de zecimale.

Cunoscînd valoarea lui α, rezultă imediat valoarea lui $\rho = x - 1$.

Această metodă, bazată pe soluții aproximative, nu poate fi satisfăcătoare. Ea nu ne poate conduce la o analiză mai amănunțită a ritmului

*) Modul de rezolvare a ecuației (I) în matematicile financiare se poate vedea în tratatul Gh. Mihoc, I. Craiu, N. Radu, D. Firescu, B. Comisioner: *Teoria matematicii în operațiile financiare*, vol. I, Ed. științifică, București, 1960, p. 161–170.

mediu al creşterii producţiei într-un interval de timp dat. A. Froda a rezolvat această dificultate, folosind în mod ingenios, credem pentru prima oară în acest deceniu, metoda continuă a ritmului de creştere.

În cele ce urmează expunem rezultatele obţinute de A. Froda prin acest procedeu.

Fie $p(t)$ producţia la momentul t. După cum se vede, presupunem că producţia este o funcţie bine determinată de timp. Admitem că $p(t)$este o funcţie continuă de t şi că admite, în intervalul considerat, derivata de primul ordin. Prin $p_0 = p(0)$ însemnăm producţia iniţială. Notăm prin $\eta(t)dt$ creşterea producţiei în intervalul de timp $(t, t + dt)$ pentru o producţie egală cu unitatea la momentul t. Creşterea $\eta(t)$ poate fi, deci, considerată ca o funcţie de t, care ne dă intensitatea de acumulare la momentul t.

Notînd prin $dp(t)$ creşterea producţiei în intervalul de timp $(t, t+dt)$, avem în baza notaţiilor de mai sus :

$$p(t) + dp(t) = p(t)\,[1 + \eta(t)dt].$$

Din această relaţie rezultă imediat :

$$\eta(t) = \frac{p'(t)}{p(t)}\,.$$

Intensitatea de acumulare este, deci, derivata logaritmică a funcţiei $p(t)$.

Urmînd aceeaşi cale ca în cazul discontinuu, vom nota prin η o intensitate medie de acumulare, care nu depinde de t. Fie $(0, T)$ intervalul de variaţie al timpului t şi $P(0, T)$ producţia totală realizată în timpul $(0, T)$.

Intensitatea medie η trebuie să fie astfel determinată ca în intervalul de timp $(0, T)$, pornindu-se de la producţia specifică p_0, să se obţină producţia totală P_{0T}.

Notăm cu $\bar{p}(t)$ producţia obţinută în baza intensităţii de acumulare η

Avem, întocmai ca şi în cazul unei intensităţi de acumulare $\eta(t)$, variabilă cu timpul, relaţia

$$\bar{p}(t) + d\bar{p}(t) = \bar{p}(t)\,[1 + \eta dt)]$$

sau

$$\frac{d\bar{p}(t)}{\bar{p}} = \eta dt.$$

Prin integrare rezultă :

$$\lg \bar{p}(t) = \int_{\alpha}^{t} \eta du$$

sau

$$\bar{p}(t) = e^{\eta(t-\alpha)}.$$

Pentru determinarea constantei α, observăm că dacă facem $t = 0$, trebuie să avem

$$\bar{p}(0) = p_0.$$

Făcînd $t = 0$, obţinem deci

$$p_0 = e^{-\eta\alpha}.$$

Introducînd valoarea lui $e^{-\eta\alpha}$ astfel obţinută în relaţia de mai sus avem

$$\bar{p}(t) = p_0 e^{\eta t}.$$

Să ţinem seama de faptul că producţia $p(t)$ acumulată în intervalul de timp $(0, T)$ trebuie să fie egală cu P_{0T}. Această producţie totală este dată de integrala :

$$\int_0^T \bar{p}(t)\, dt$$

Înlocuind în această integrală pe $\bar{p}(t)$ prin valoarea găsită mai sus, căpătăm :

$$\int_0^T \bar{p}(t)\,dt = \int_0^T p_0 e^{\eta t}\, dt = \frac{P_0}{\eta}[e^{\eta t}]_0^T = p_0 \frac{e^{\eta T} - 1}{\eta}.$$

Ajungem deci la ecuaţia finală

$$p_0 \frac{e^{\eta T} - 1}{\eta} = P_{0T}.$$

Introducînd notaţia

$$r = \frac{P_{0T}}{p_0 T}$$

ajungem la ecuaţia

$$\text{(III)} \qquad \frac{e^{\eta T} - 1}{\eta T} = r.$$

Aceasta este relaţia fundamentală care leagă între ele numerele η, T, r.

Desigur că la această relaţie se putea ajunge şi direct, fără intermediul unei ecuaţii diferenţiale.

Astfel producţia totală P_{0t}, în intervalul de timp $(0, t)$ este

$$P_{0t} = \int_0^t p(t)dt.$$

În intervalul $(0, T)$ avem deci

$$P_{0T} = \int_0^T p(t)dt.$$

Producția specifică mijlocie p_{0T} este

$$p_{0T} = \frac{P_{0T}}{T} = \frac{1}{T}\int_0^T p(t)\mathrm{d}t.$$

Împărțim intervalul $(0, T)$ în $n+1$ intervale egale cu Δt:

$$T = (n+1)\,\Delta t.$$

În formula (II) introducem notația

$$\rho = \eta\,\Delta t;$$

obținem

$$\frac{(1+\eta\Delta t)^{n+1}-1}{\eta\Delta t} = \frac{P_{0T}}{\bar{p}_0\Delta t} = r.$$

Sau, dacă ținem seama de faptul că

$$n+1 = \frac{T}{\Delta t},$$

$$\frac{(1+\eta\Delta t)^{\frac{T}{\Delta t}}-1}{\eta T} = r.$$

Pentru $\Delta t \to 0$, relația de mai sus ne dă

$$\text{(IV)} \qquad \frac{e^{\eta T}-1}{\eta T} = r$$

adică egalitatea (III), pe care am obținut-o și pe altă cale.

În formula de mai sus avem întotdeuna $T > 0$ de aici rezultă inegalitățile respective:

$$\begin{array}{lll} \eta > 0\,, & \eta = 0\,, & \eta < 0, \\ r > 1\,, & r = 1\,, & r < 1. \end{array} \qquad \text{(XIV,4.3')}$$

Vom observa că în practică nu ne interesează valoarea lui η, ci valoarea ritmului mediu anual ρ, cînd ni se dau datele producției, concretizate prin r. Este mai greu ca direct, din relația (II), să se găsească unele delimitări pentru ρ. În funcție de η este însă posibil ca să se realizeze acest lucru.

În relația (IV) unitatea de timp nu este specificată.

Să punem $T=(n+1)t$, ceea ce revine la a nota prin t durata unui an. Ajungem, ținînd seama de (II) și (IV), la

$$\frac{(1+\rho)^{n+1}-1}{(n+1)\rho} = \frac{e^{\eta T}-1}{\eta T} = r.$$

Alegem acum anul ca unitate de măsură: $t = 1$. Relaţiile precedente devin

$$\text{(V)} \qquad \frac{(1+\rho)^{n+1}-1}{\rho} = \frac{e^{(n+1)\eta}-1}{\eta} = (n+1)\,r.$$

Am obţinut astfel două relaţii simple între ρ, η şi r.

În baza analizei pînă acum făcute am putut stabili inegalitatea dublă

$$|\sqrt[n]{r}-1| < |\rho| < |\sqrt[n]{(n+1)r}-1|,$$

care ne dă două delimitări pentru ρ.

Aceste delimitări sînt însă prea mari. A. Froda a dovedit că folosirea relaţiilor (V) ne poate conduce la delimitări mai strînse.

Inegalităţile stabilite de Froda sînt următoarele:

$$\text{(VI)} \begin{cases} e^{\eta}-1 < \rho < \dfrac{2}{n}(r-1) \quad \text{pentru } r>1, \quad \rho>0, \\ -\dfrac{3}{2(n-1)} + \sqrt{\dfrac{9}{4(n-1)^2} - \dfrac{6(1-r)}{n(n-1)}} < \rho < e^{\eta}-1, \text{ pentru } r<1, \ \rho<0. \end{cases}$$

După cum vom vedea, aceste inegalităţi conduc la intervale de variaţie mai strînse decît cele mai sus arătate.

Mai este de observat că în inegalităţile (VI) intră η. Aceasta nu este însă un inpediment, deoarece η se exprimă uşor în funcţie de r. Într-adevăr, în relaţia

$$\frac{e^{\eta T}-1}{\eta T} = r$$

facem substituţia

$$\eta T = z.$$

Rezultă

$$\frac{e^{z}-1}{z} = r.$$

Din această ecuaţie se poate determina z în funcţie de r. În acest scop se poate întocmi un tabel care să dea pentru fiecare valoare a lui r, valoarea corespunzătoare a lui z, cu un număr oarecare de zecimale exacte.

Dacă durata T este egală cu $(n+1)$ ani, rezultă

$$\eta = \frac{z}{n+1}.$$

Putem, deci, considera, din punct de vedere practic, pe η cunoscut pentru fiecare valoare dată a lui r.

Din această cauză inegalităţile (VI) au un caracter practic, fiind utilizabile în diverse aplicaţii.

Vom demonstra acum inegalitățile (VI).

Notăm:

$$\rho_1 = e^{\eta} - 1$$

Presupunem mai întîi

$$\rho > 0, \ \eta > 0.$$

În acest caz putem scrie:

$$\frac{(1+\rho_1)^{n+1}-1}{\rho_1} = \frac{e^{(n+1)\eta}-1}{\rho_1} < \frac{e^{(n+1)\eta}-1}{\eta} = \frac{(1+\rho)^{n+1}-1}{\rho},$$

deoarece

$$e^{\eta} - 1 = \left(1 + \eta + \frac{\eta^2}{2} + \ldots\right) - 1 > (1+\eta) - 1 = \eta > 0.$$

Să mai notăm

$$R(y) = \frac{(1+y)^{n+1}-1}{y} = 1 + (1+y) + (1+y)^2 + \ldots + (1+y)^n.$$

Ultima dezvoltare ne arată că $R(y)$ este o funcție crescătoare în raport cu y.

Cum am stabilit că

$$R(\rho_1) < R(\rho),$$

reiese că

$$\rho_1 < \rho.$$

Am demonstrat prin urmare că

$$\rho > e^{\eta} - 1,$$

adică una din inegalitățile (VI).

Examinăm acum cazul $\rho < 0$, $\eta < 0$.

Vom proceda în același mod, folosind notația

$$\bar{\rho} = -\rho.$$

Vom putea scrie în acest caz

$$\frac{1-(1-\bar{\rho}_1)^{n+1}}{\bar{\rho}_1} = \frac{1-[1-(1-e^{\eta})]^{n+1}}{1-e^{\eta}} > \frac{1-e^{(n+1)\eta}}{-\eta} = \frac{1-(1-\bar{\rho})^{n+1}}{\bar{\rho}}.$$

Inegalitatea se justifică prin faptul că

$$0 < 1 - e^{\eta} = 1 - \left(1 + \eta + \frac{\eta^2}{2!} + \ldots\right) < 1 - (1+\eta) = -\eta.$$

Vom arăta în mod analog

$$\bar{R}(y) = \frac{1-(1-y)^{n+1}}{y} = 1 + (1-y) + (1-y)^2 + \ldots + (1-y)^n.$$

Funcția $R(y)$ descrește cînd y crește.

Cum am demonstrat că

$$\bar{R}(\bar{\rho}_1) > \bar{R}(\bar{\rho}),$$

rezultă, din proprietatea sus-menționată a funcției $\bar{R}(y)$, că avem

$$\bar{\rho}_1 < \bar{\rho}$$

sau

$$\rho < e^{\eta} - 1.$$

Am demonstrat astfel încă una din inegalitățile (VI).

Revenim, în continuare, la cazul $\rho > 0$, $\eta > 0$. Notăm

$$\rho_2 = \alpha\eta,$$

α fiind un număr pozitiv, pe care vrem să-l determinăm astfel încît să avem

$$\rho < \rho_2$$

Ținînd seama de (V) și de faptul $R(\rho_2) > R(\rho)$, obținem

$$\frac{(1+\rho_2)^{n+1}-1}{\rho_2} = \frac{(1+\alpha\eta)^{n+1}1}{\alpha\eta} > \frac{e^{(n+1)\eta}1}{\eta} = \frac{(1+\rho)^{n+1}-1}{\rho}.$$

De aici rezultă că:

$$(1+\alpha\eta)^{n+1} - 1 > \alpha[e^{(n+1)\eta} - 1].$$

Din dezvoltarea lui $(1+\alpha\eta)^{n+1}$ păstrăm numai primii trei termeni și pentru ca inegalitatea să fie satisfăcută este suficient să avem

$$(n+1)\alpha\eta + \frac{n(n+1)}{2}\alpha^2\eta^2 = \alpha e^{(n+1)\eta} - \alpha.$$

Din această relație rezultă

$$\rho_2 = \frac{2}{n}\left[\frac{e^{(n+1)\eta}-1}{(n+1)\eta} - 1\right] = \frac{2}{n}(r-1),$$

sau

$$\rho < \frac{2}{n}(r-1).$$

Am demonstrat astfel încă una din inegalitățile (VI).

Pentru a dovedi exactitatea inegalității, care ne-a mai rămas, examinăm din nou cazul $\rho < 0$, $\eta < 0$.

La fel, însemnăm $\bar{\rho} = -\rho$ și ne propunem să determinăm pe $\bar{\alpha}$, astfel ca

$$\bar{\rho}_2 = -\bar{\alpha}\eta, \qquad \bar{\alpha} > 0, \qquad \bar{\rho} < \bar{\rho}_2.$$

Deoarece din ultima inegalitate rezultă

$$\bar{R}(\bar{\rho}) > \bar{R}(\bar{\rho}_2),$$

deducem din (V)

$$\frac{1-(1-\bar{\rho}_2)^{n+1}}{\bar{\rho}_2} = \frac{1-(1+\bar{\alpha}\eta)^{n+1}}{-\bar{\alpha}\eta} < \frac{1-e^{(n+1)\eta}}{-\eta} = \frac{1-(1-\bar{\rho})^{n+1}}{\rho}.$$

Din această inegalitate rezultă, în condițiile noastre:

$$1-(1+\bar{\alpha}\eta)^{n+1} < \bar{\alpha}[1-e^{(n+1)\eta}].$$

Dacă presupunem

$$\bar{\rho}_2 < \frac{5}{n-3}, \qquad (n>3)$$

se poate dovedi uşor că pentru ca inegalitatea precedentă să fie satisfăcută este suficient ca

$$\frac{n(n-1)}{6}\,\bar{\alpha}^2\eta^2+\frac{n}{2}\,\bar{\alpha}\eta+1=\frac{1-e^{(n+1)\eta}}{(n+1)\eta}=r$$

sau

$$\bar{\rho}_2^2-\frac{3}{n-1}\,\bar{\rho}_2+\frac{6}{n(n-1)}\,(1-r)=0.$$

Condiţia ca rădăcinile acestei ecuaţii să fie reale este

$$\frac{9}{(n-1)^2}-\frac{24(1-r)}{n(n-1)}\geqslant 0.$$

De aici deducem

$$r\geqslant\frac{5n-8}{8(n-1)}.$$

Admitem această ipoteză, care este suficientă pentru aplicaţiile practice.

Cea mai mică rădăcină reală $\bar{\rho}_2$ este

$$\bar{\rho}_2=\frac{3}{2(n-1)}-\sqrt{\frac{9}{4(n-1)^2}-\frac{6(1-r)}{n-1}}.$$

Cum $\bar{\rho}<\bar{\rho}_2$, rezultă $\rho>-\bar{\rho}_2$ adică

$$\rho>-\frac{3}{2(n-1)}+\sqrt{\frac{9}{4(n-1)^2}-\frac{6(1-r)}{n(n-1)}}.$$

Am demonstrat astfel toate inegalităţile.

5. Metoda tendinţei exponenţiale

I. Pisarev, P. Maslov*), P. Köves, I. Marinescu şi P. Petrov recomandă ca ritmul mediu să fie calculat pe baza tendinţei centrale de tip exponenţial, folosind metoda de ajustare analitică a celor mai mici pătrate.

Din punct de vedere teoretic, aplicarea acestei metode se justifică prin existenţa unei creşteri relativ uniforme a valorilor unei serii, care poate fi exprimată printr-o funcţie exponenţială de forma

$$y=ab^t$$

în care, după cum ştim, y reprezintă valorile calculate ale seriei, t timpul, iar a şi b constante. Constanta a indică punctul de intersecţie a curbei cu axa ordonatelor, iar b gradul de înclinaţie a curbei — panta curbei sau viteza ei aşa cum s-a mai arătat la (XIII, 4.4) — adică tocmai *ritmul mediu* de creştere.

Nu vom reveni aici asupra modului de calcul al parametrilor a şi b prin metoda celor mai mici pătrate.

*) P. Maslov, *Analiza critică a publicaţiilor statisticii burgheze*. Ed. ştiinţifică, Bucureşti, 1957.

Remarcăm şi aici o observaţie interesantă a lui P. Petrov. El arată că în cazul cînd modelul matematic al seriei studiate este de tipul progresiei geometrice, exprimat în formă generală printr-o funcţie exponenţială, prima derivată a acestei funcţii este o mărime variabilă iar raportul dintre această derivată şi funcţia respectivă este o mărime constantă, egală cu lg b. De aici rezultă că dacă schimbările unei caracteristici cantitative, în funcţie de timp, au ca model o progresie geometrică, ritmul mediu de creştere se va apropia de acest ritm constant, existent în realitate.

Este evident că numai într-un singur caz este posibil un ritm constant de creştere — atunci cînd modelul schimbării unei anumite caracteristici cantitative este o progresie geometrică. Un asemenea model se poate întîlni de obicei cînd se schimbă numărul cazurilor unui fenomen şi cînd fenomenul se caracterizează prin capacitatea de înmulţire şi reproducţie — populaţia, de pildă — precum şi în unele situaţii mai rare şi anume, cînd schimbarea unor caracteristici cantitative urmează modelul unui produs sau raport dintr-o funcţie exponenţială şi o constantă.

6. Cîteva concluzii asupra procedeelor de calcul al ritmului mediu de creştere

1°. Este incontestabil că pe calea mediei geometrice, în ultimă analiză, are loc o nivelare specială a unei serii dinamice, asemănătoare ajustării făcute prin metoda celor mai mici pătrate, cu deosebirea esenţială că ea nu ţine seama de toate mărimile seriei, ci are numai două puncte de sprijin — prima şi ultima mărime a seriei, aşa cum rezultă din formula (XIV, 2.3) care se foloseşte de obicei în practică. Această particularitate limitează, evident, sau face să fie nesigură, folosirea mediei geometrice, chiar şi în cazurile cînd analiza obiectivă a fenomenului studiat ne-a dezvăluit suficiente temeiuri de a calcula ritmul mediu de creştere. Aceste restricţii provin din faptul că cei doi termeni consideraţi în calcul pot fi pe drept cuvînt influenţaţi de acţiunea unor factori întîmplători, din care cauză poate să ne apară o serie dinamică ajustată care să nu aibă nimic comun cu seria reală, în afara primului şi ultimului termen.

Cel mai potrivit şi sigur domeniu de aplicare a mediei geometrice este acela al calculului ritmului mediu de creştere în planurile de perspectivă. Cînd se planifică schimbările privind un fenomen oarecare care se exprimă prin schimbările respective ale caracteristicii şi se ştie că aceste schimbări vor avea, ca model matematic, forma unei progresii geometrice, atunci este de ajuns să se dea cele două puncte de sprijin — nivelul de la începutul şi sfîrşitul perioadei de plan — pentru a putea, cu ajutorul mediei geometrice calculate, să se obţină toată seria de indicatori de plan care trebuie realizată în procesul de îndeplinire a planului. În afară de aceasta, cele două valori de plan — prima ca un nivel real de la care se porneşte, iar a doua ca un indicator planificat — nu pot fi socotite ca mărimi întîmplătoare, care poartă în sine influenţa factorilor întîmplători.

Scanned with OKEN Scanner

Rezultă deci, că singurul criteriu corect care ne poate ajuta la alegerea mediei geometrice pentru calculul ritmului mediu este *considerarea particularităților concrete în care se dezvoltă fenomenul cercetat*, gradul lui de apropiere de o progresie geometrică.

Este știut, de pildă, că în practica statisticii socialiste, ritmul mediu de creștere a populației se calculează cu ajutorul mediei geometrice. Acest procedeu nu este greșit întrucît el este justificat de *procesul material de reproducere a populației* *). În afară de aceasta, calculul mediei geometrice este indicat și din cauză că, de obicei, se cunosc numai primul și ultimul termen al seriei (datele a două recensăminturi periodice).

Media geometrică este folosită și la noi la calculul ritmului mediu planificat, ceea ce, după cum am văzut, este justificat din punct de vedere teoretic întrucît la elaborarea planului se ține seama de legile interne specifice de dezvoltare ale fenomenelor respective.

2°. Calculul ritmului mediu prin metoda mediei parabolice are într-adevăr deficiențele semnalate de diverși statisticieni, arătați mai sus; ele decurg din faptul că, în esență, în calculul ritmului mediu de dezvoltare trebuie să ținem seama numai de valoarea primului termen și de suma termenilor.

Trebuie să arătăm însă că o lipsă vizibilă a metodei mediei parabolice, și care este proprie de altfel și mediei geometrice, constă în faptul că ea desprinde din realitatea complexă numai un număr mic de faze. Nici media geometrică, nici media parabolică nu sînt utilizabile dacă seria dinamică nu prezintă nici un fel de dezvoltare uniformă, în sens absolut sau relativ.

Prin urmare, ori de cîte ori sîntem în fața unor serii dinamice cu o *anumită constanță relativă* a creșterii (descreșterii) — pentru care, deci, cercetarea ritmului mediu de dezvoltare este pe deplin justificată — ambele metode (media geometrică și media parabolică) pot fi utilizate deopotrivă, în linii mari ele fiind echivalente.

Să luăm exemplul din tabelul 60.

Tabelul 60

Dinamica producției de construcții-montaj în R.P.R., în sectorul socialist, pe perioada 1955—1959

Ani	1955	1956	1957	1958	1959
Indici de bază	100	115	101	104	111

Sursa : *Anuarul statistic al R.P.R.*, 1959 și Buletinul statistic trimestrial, nr. 4 (1959).

a) Calculînd ritmul mediu prin metoda mediei geometrice obținem :

$$\bar{\theta}_g = \sqrt[4]{\frac{111}{100}} = \sqrt[4]{1{,}11}$$

*) Asupra numărului populației influențează evident și alte cauze, însă principalul constă în aceea că aceste cauze nu pot nici să reducă capacitatea populației de a se reproduce, nici să ducă la schimbarea arbitrară a numărului acesteia (cu excepția unor cazuri particulare, foarte rare).

care ne dă

$$4^{-1} \lg 1{,}11 = 0{,}04532 \cdot 4^{-1} = 0{,}01133.$$

de unde rezultă că

$$\bar{\theta}_g = 1{,}0264.$$

b) Calculul ritmului mediu prin metoda mediei parabolice revine la a găsi rădăcina cuprinsă între 1 şi 2 a ecuaţiei (XIV, 3.3). Avem

$$\frac{\theta^5 - 1}{\bar{\theta} - 1} = \frac{531}{100} = 5{,}31,$$

de unde

$$\bar{\theta} \approx 1{,}03.$$

După cum se vede, nivelul ritmului mediu calculat prin cele două metode nu prezintă deosebiri semnificative.

3°. Calculul ritmului mediu, folosind ajustarea unei drepte sau curbe exponenţiale prin metoda celor mai mici pătrate, este cel mai indicat.

Într-adevăr, la urmărirea dinamicii şi ritmului mediu de dezvoltare a fenomenului se caută o anumită continuitate şi se presupune în acest scop că în perioada cercetată a existat o linie permanentă în ceea ce priveşte schimbarea şi care este exprimată printr-un ritm constant oarecare, a cărui mărime aproximativă se va obţine prin calculul mediei. În ipoteza în care nu se consideră aceste condiţii, calculul mediei nu are un sens real. Dacă pentru intensitatea schimbării pe care o studiem nu este caracteristică o asemenea continuitate ci dimpotrivă un ritm neregulat care se modifică mereu, atunci ritmul mediu calculat va deforma realitatea.

În ultimă analiză, deci, cercetarea ritmului mediu de creştere înseamnă aproximarea evoluţiei reale a seriei dinamice printr-o schemă teoretică oarecare, printr-un model teoretic.

Or, criteriul aproximaţiei poate fi ales mai bine prin calculul tendinţei centrale a fenomenului studiat, prin metoda celor mai mici pătrate.

În legătură cu problema alegerii tendinţei centrale liniare sau exponenţiale ea se rezolvă, în general, pe baza criteriilor şi elementelor pe care le-am dat în capitolul precedent la alegerea însăşi a tipului de curbă care aproximează tendinţa centrală.

Scanned with OKEN Scanner

PARTEA A PATRA

METODA INDICILOR

Scanned with OKEN Scanner

Scanned with OKEN Scanner

Capitolul XV

Esenţa indicilor. Clasificarea lor şi metodele de calcul

1. Noţiuni generale

În teoria şi practica statisticii socialiste, prin categoria statistică a indicilor se înţelege o expresie de un anumit fel a mărimilor relative, avînd funcţia de a măsura, în primul rînd, *schimbarea medie în timp* — aspect dinamic — a fenomenelor şi proceselor social-economice sau de a le *compara în spaţiu*, la un moment dat sau într-o anumită perioadă, aspect static *).

Indicii se mai folosesc pentru a caracteriza gradul de îndeplinire a sarcinilor de plan sau pentru a reflecta, la întocmirea planului, însuşi nivelul relativ al acestor sarcini.

Procedeul de construire a oricărui indice rezultă dintr-o relaţie matematică elementară: se construieşte un raport prin care anumite date statistice luate în studiu se compară cu alte date, cu caracter analog, dintr-o altă perioadă sau din aceeaşi perioadă, însă, din spaţiu diferit. Datele luate în studiu — *date raportate* — apar la numărătorul raportului, iar datele cu care se face comparaţia — *bază de raportare* — apar la numitorul acestuia.

Pe cît de simplă este schema matematică a oricărui indice, pe atît de complexe sînt, pe de altă parte, conţinutul şi metodologia de calcul al indicelui, ca procedeu de analiză statistică a fenomenelor şi proceselor social-economice, deci, ca *metodă specifică* a statisticii în înţelesul ei clasic de ştiinţă socială, fundamentată pe tezele materialismului istoric şi ale economiei politice marxist-leniniste.

*) Vezi în acest sens: *Teoria statisticii*. Editura de stat pentru literatură economică şi juridică, Bucureşti, 1955, p. 164.

De aceea, problema centrală care se pune atunci cînd elaborăm un indice economic constă în a înțelege conținutul, *esența* fenomenului economic pentru care se calculează. Fără rezolvarea acestei probleme primordiale, construirea oricărui indice economic este de neconceput.

Faptul că prin metoda indicilor sîntem în măsură să caracterizăm statistic esența fenomenelor studiate are un înțeles teoretic precis. În general, datele luate în studiu (cele raportate, care figurează la numărătorul raportului), caracterizează un aspect static al fenomenului respectiv. Dar, de îndată ce le apreciem comparativ, în raport cu datele dintr-o perioadă aleasă ca bază și care reprezintă o altă etapă a dezvoltării fenomenului studiat, noi relevăm aspectul dinamic al dezvoltării, mișcarea și schimbarea fenomenului în timp. Tocmai în aceasta constă sensul principal și semnificația comparării prin metoda indicilor.

Compararea datelor curente cu datele perioadei de bază ne oferă posibilitatea de a sesiza noul pe care-l cuprind, de a măsura și caracteriza proporțiile schimbărilor care au survenit și nivelul de dezvoltare atins, precum și de a releva *calitatea* specifică a fenomenului studiat.

Ca metodă de analiză economico-statistică, metoda indicilor are o serie de trăsături caracteristice care o deosebesc de celelalte metode statistice, o apropie însă, foarte mult, de metoda leninistă a grupărilor statistice.

Trebuie să atragem atenția că adeseori în practica statisticii se manifestă tendința de a rezolva — prin calculul unui singur indice — probleme deosebit de complexe, depășind posibilitățile de caracterizare a indicelui construit. Are loc — după cum se spune — o *supraîncărcare* a sarcinilor, a funcțiilor și a rolului pe care un indice este chemat a le îndeplini și, în consecință, devine dificilă înțelegerea semnificației lui reale, interpretarea lui corectă.

Din punctul de vedere al sferei de cuprindere distingem două categorii de indici: *indici individuali*, exprimînd raportul de mărime — în timp sau spațiu — al unui fenomen (element) oarecare dintr-o colectivitate statistică și *indici generali* sau *de grup*, exprimînd schimbări medii ale întregii colectivități de fenomene (obiecte) sau ale caracteristicilor unor grupe de fenomene — părți ale colectivității — supuse studiului.

Se pot da numeroase exemple de indici individuali exprimînd raportul cantitativ al producției diferitelor produse — în perioada curentă față de perioada de bază — exprimată în aceleași unități de măsură.

De exemplu, în Documentele istorice ale Congresului al III-lea al P.M.R., cuprinzînd directivele pentru dezvoltarea economiei naționale pe anii 1960—1965 ca și în schița-program de dezvoltare a economiei naționale, în perspectivă, pînă în anul 1975, sarcinile principale sînt date sub forma indicilor individuali.

Se prevede astfel că, pînă în 1965, producția de energie electrică va crește de aproximativ 2,7 ori, iar pînă în 1975 de aproximativ 10 ori, de la 6,8 miliarde kW în 1959 la 65—70 miliarde kW, ceea ce va asigura industriei mai multă energie electrică și va permite introducerea, pe scară tot mai largă, a proceselor tehnologice moderne, electrificarea agriculturii, a căilor ferate, precum și a orașelor și satelor noastre. Sau,

Scanned with OKEN Scanner

producția de oțel, pînă în 1965, va crește de 2,3 ori față de 1959, iar a minereurilor de fier de 3,8 ori etc.

De asemenea, din aceleași Documente ale Congresului al III-lea al P.M.R. se pot da exemple de indici generali — de grup. Ne referim, de exemplu, la valoarea producției globale industriale, care în anul 1965 va fi de 2,1 ori mai mare decît în 1959, corespunzînd cu un ritm mediu anual de creștere de aproximativ 13%.

Din definiția indicelui care a fost dată mai sus, ca și din exemplele considerate, rezultă clar că indicele reprezintă tot o mărime *relativă* avînd anumite particularități care îl deosebesc de celelalte mărimi relative. Aceste particularități apar îndeosebi atunci cînd se calculează indici generali, de grup, sau indici *analitici*, caracterizînd dinamica medie a unor fenomene complexe cu două sau mai multe caracteristici, neînsumabile direct. Tocmai de aceea, noțiunea de *indice* se referă — în sensul ei strict tehnic — la indice de grup *).

La întocmirea indicilor de grup se pun o serie de probleme legate de măsurarea obiectelor luate în studiu.

Este evident, de exemplu, că pentru a compara dinamic sau sincronic producția industrială este necesar, în prealabil, să se rezolve corect metoda de determinare a volumului producției. Or, volumul total al producției industriale este un fenomen complex, în compunerea căruia intră o serie de elemente (o tonă de cărbune, o tonă de zahăr etc.) ce nu se pot măsura cu aceeași unitate de măsură.

La elaborarea indicelui se pune, de aceea, problema întrunirii lor într-un singur volum, adoptîndu-se în acest scop, drept unitate de măsură, o caracteristică ce le este comună. În teoria și practica statisticii socialiste, valoarea diferitelor produse — îndeosebi valoarea lor în expresie bănească — a fost adoptată ca unitate de măsură comună.

2. Indici în formă de bază și în formă de lanț

Una din problemele principale ce se pun la construirea indicilor este problema *bazei de comparație*. Adeseori, în lucrările practice de statistică, drept bază de comparație, se ia sarcina de plan, știut fiind că urmărirea sistematică a cursului îndeplinirii planului de dezvoltare a economiei naționale este una din sarcinile fundamentale ale statisticii din țara noastră. Pe de altă parte, am văzut că înseși sarcinile de plan sînt formulate sub forma indicilor. Pentru a urmări, de exemplu, sarcina planificată de creștere a productivității muncii în perioada 1960—1965, în siderurgie cu 50%, în industria constructoare de mașini cu 75—80%, în industria chimică cu aproximativ 80% etc., sau pe ansamblul industriei, în aceeași perioadă, cu 60—65%, se impune cu necesitate ca statisticianul să calculeze indicii statistici respectivi. În acest scop, el trebuie să considere

*) După unii autori, noțiunea de „indici" trebuie să se refere atît la indicii individuali (sau indici *simpli*) cît și la indicii *analitici*. Ambele feluri de indici reprezintă categorii ale uneia și aceleiași clase a indicatorilor statistici — clasa indicilor.

datele reale caracterizînd nivelul productivității muncii, atins în fiecare an al şesenalului, şi să le compare cu datele privind productivitatea muncii din anul luat drept bază de comparaţie — acelaşi care a fost considerat şi în planul de perspectivă (anul 1959). Comparînd indicatorii obţinuţi pe baza datelor efective cu indicatorii de plan el va putea caracteriza gradul de îndeplinire a planului, de creştere a productivităţii muncii.

Construirea indicelui, luînd drept bază de comparaţie o perioadă anterioară, prezintă interes nu numai din punctul de vedere al asigurării comparabilităţii cu datele de plan. Problema ritmului de dezvoltare a economiei naţionale, a diferitelor ei ramuri, a culturii socialiste, reprezintă prin ea însăşi un interes deosebit.

În legătură cu aceasta este suficient să amintim interesul uriaş şi entuziasmul cu care a fost primit Raportul C.C. al P.M.R. la cel de-al III-lea Congres al partidului, în care au fost folosite din plin şi metodele de analiză ale ştiinţei statistice şi, îndeosebi, metoda indicilor *), pentru a compara realizările oamenilor muncii din perioada 1956—1959 cu cele din anul 1955 sau din anii anteriori, luaţi drept bază de comparaţie.

Pentru construirea indicilor se mai poate lua şi o altă bază de comparaţie şi anume — nivelul aceluiaşi fenomen din alt loc al spaţiului (din altă întreprindere, regiune etc.), care pune probleme specifice ale comparării pe plan interregional sau internaţional. Vom acorda acestor probleme o atenţie deosebită, în cadrul capitolului XVI privind indicii teritoriali.

În funcţie de perioada aleasă drept bază de comparaţie distingem două feluri de indici: *indici în formă de bază*, cînd compararea datelor pe diferite perioade se face în raport cu una şi aceeaşi perioadă de bază şi *indici în formă de lanţ* (cu bază variabilă), cînd compararea datelor privind nivelul unor indicatori economici pe diferite perioade se face în raport cu datele corespunzătoare ale perioadei precedente.

Dacă notăm, de exemplu, cantităţile unui anumit produs, fabricate în 10 ani succesivi prin

$$q_0,\ q_1,\ q_2,\ q_3,\ q_4,\ \ldots,\ q_9,$$

atunci seria indicilor în formă de bază este:

$$\frac{q_1}{q_0},\ \frac{q_2}{q_0},\ \frac{q_3}{q_0},\ \ldots,\ \frac{q_9}{q_0}$$

iar a indicilor în formă de lanţ

$$\frac{q_1}{q_0},\ \frac{q_2}{q_1},\ \frac{q_3}{q_2},\ \ldots,\ \frac{q_9}{q_8}.$$

În acest exemplu este vorba de indici individuali exprimaţi în aceleaşi unităţi de măsură.

O proprietate remarcabilă a indicilor individuali în formă de lanţ constă în aceea că din serii de asemenea indici putem construi serii cores-

*) Gh. Gheorghiu-Dej, *Raport la cel de al III-lea Congres al Partidului Muncitoresc Romîn*. Ed. politică, Bucureşti, 1960, p. 66.

punzătoare de indici de bază. Acest lucru se obține făcînd produsul succesiv al indicilor în lanț.

De exemplu, dacă considerăm produsul indicilor în formă de lanț pe primii cinci ani, folosind notația de mai sus, obținem indicele de bază al celui de-al cincilea an, adică raportul:

$$i_{4/0} = \frac{q_4}{q_0}.$$

Pentru indicii generali, de grup, această proprietate se menține numai dacă este asigurată pe deplin comparabilitatea datelor, folosindu-se ponderi constante la elaborarea indicilor de grup în lanț. În continuare, vom avea prilejul să arătăm, concret, cazul cînd această condiție este îndeplinită.

Legătura dintre indicii în formă de lanț și indicii în formă de bază se poate exprima și sub o altă formă — inversă și anume: împărțind indicele în formă de bază al perioadei k la indicele în formă de bază al perioadei $k-1$, se obține indicele în formă de lanț al perioadei k, față de perioada $k-1$.

Astfel, considerînd indicii în formă de bază ai ultimilor patru ani din exemplul de mai sus și făcînd rapoartele respective, în sensul arătat, obținem șirul respectiv al indicilor în formă de lanț:

$$\frac{q_7}{q_0} : \frac{q_6}{q_0} = \frac{q_7}{q_6}; \quad \frac{q_8}{q_0} : \frac{q_7}{q_0} = \frac{q_8}{q_7}; \quad \frac{q_9}{q_0} : \frac{q_8}{q_0} = \frac{q_9}{q_8}.$$

Legătura dintre indicii în formă de bază și indicii în formă de lanț a mai fost scoasă în evidență și în capitolul XII (vezi p. 300).

Indicii în formă de bază sînt instrumente prețioase pentru a caracteriza rezultatele muncii economice pe perioade de timp mai mari, spre deosebire de indicii în formă de lanț care ilustrează schimbarea nivelului unui fenomen sau proces economic în perioada considerată, față de cea precedentă.

Să ilustrăm semnificația indicilor în formă de bază și formă de lanț printr-un exemplu concret (vezi tabelul 61).

Tabelul 61

Dinamica productivității muncii în unități naturale la unele produse industriale ale R.P.R.

Indicatori	1956	1957	1958	1959	1956	1957	1958	1959
	indici în formă de bază				indici în formă de lanț *)			
0	1	2	3	4	5	6	7	8
Producția de energie electrică pe un muncitor	113	140	155	167	113	124,0	110,7	107,7
Producția de cărbune pe un muncitor din ramura cărbune	105	111	119	130	105	105,7	107,2	109,2
Producția de țiței pe un muncitor ocupat în extracție	104	107	109	111	104	102,8	102,8	101,8
Producția de ciment pe un muncitor ocupat în fabricile de ciment	112	123	132	145	112	109,8	107,3	109,8

Notă: Date luate din „Breviarul Statistic al R.P.R., 1960".
*) Indici calculați pe baza datelor din primele patru coloane.

Scanned with OKEN Scanner

Examinînd datele din acest tabel ne convingem uşor că fiecare din seriile indicilor individuali — de bază şi în lanţ — are un sens anumit : indicii de bază ilustrează realizările obţinute în decursul întregii perioade ce s-a scurs de la anul luat drept bază de comparaţie, iar indicii în lanţ — rezultatele fiecărui an în parte, faţă de cele din anul precedent. Aşa, de pildă, indicele în formă de bază în anul 1958 la producţia de energie electrică pe un muncitor ne arată că în trei ani (1956, 1957 şi 1958) nivelul acestui indicator a crescut cu 55%, în comparaţie cu anul de bază 1955.

3. Clasificarea indicilor după funcţiile lor cognitive. Schema lui V. Peregudov*)

În dezvoltarea ei continuă, teoria statisticii socialiste a elaborat un ansamblu de principii călăuzitoare pentru construirea indicilor generali, de grup. Aceasta a avut loc sub influenţa cerinţelor mereu noi ale practicii construirii socialismului şi comunismului. Trebuie arătat însă că tocmai la construirea indicilor de grup se manifestă particularităţile specifice ale metodei indicilor, dificultăţile metodologice de calcul. Se impune, de aceea, sistematizarea acestor principii şi considerarea lor în lumina noilor probleme teoretice relevate de cercetătorii sovietici şi din alte ţări socialiste. Vom apela în acest scop, îndeosebi, la consideraţiile teoretice interesante ale cunoscutului statistician şi economist sovietic V. Peregudov.

În interesul generalizării şi al clarităţii expunerii vom folosi şi notaţia corespunzătoare, introdusă de V. Peregudov.

Indicatorii economici *calitativi* (productivitatea muncii, preţul de cost etc.) vor fi simbolizaţi prin α, β,..., literele iniţiale ale alfabetului grecesc, iar cei *cantitativi*, prin ω, ψ, ... literele lui finale.

De obicei, un fenomen complex supus studiului are un dublu caracter : el apare sub forma produsului a doi indicatori din care unul este un indicator de volum (cantitativ) pe care-l notăm cu ψ, iar celălalt — un indicator de calitate (calitativ) pe care-l notăm cu α (sau β, ...) conform convenţiei de mai sus.

Ambii indicatori pot fi consideraţi ca *elemente* ale fenomenului complex, iar acesta din urmă, ca indicator funcţional **).

De exemplu, volumul circulaţiei mărfurilor (ω) poate fi exprimat prin următoarea relaţie :

$$\omega = pq$$

în care

$p = \alpha$ reprezintă preţul — element *calitativ* ;
$q = \psi$ volumul producţiei desfăcute — element *cantitativ* ;
ω este în acest caz un indicator funcţional.

*) Vezi V. N. Peregudov, *Ucionîie Zapiski po statistike*. vol. I, Moskva, 1955.

**) Prin „indicator funcţional", aici şi mai departe, se înţelege expresia unui *sistem* de caracteristici, determinat în mod obiectiv pe baza unei analize calitative multilaterale.

Complexitatea unui fenomen poate fi apreciată și dintr-un alt punct de vedere și anume, dacă este vorba de o masă de obiecte eterogene, în care caz caracteristica lui generală este determinată de mărimea însemnată cu

$$\Sigma\omega = \Sigma\alpha\psi.$$

În cazul particular, cînd este supus studiului un singur obiect omogen, lipsește semnul sumei

$$\omega = \alpha\psi.$$

În diferite probleme concrete elementul funcțional poate fi, în toate cazurile, însumat.

În ceea ce privește elementul cantitativ ψ se pot ivi trei situații:

a) Acest indicator nu se poate însuma. Din punct de vedere economic nu există $\Sigma\psi$. De exemplu nu se poate însuma volumul diferitelor produse, în forma lor specifică naturală;

b) Acest indicator se poate însuma fiind vorba, de exemplu, de producția omogenă fabricată la mai multe întreprinderi. Această însumare se notează cu $\Sigma'\psi$;

c) Sînt posibile însumări parțiale $\Sigma'\psi$ însă imposibile însumările $\Sigma\psi$ și $\Sigma\Sigma'\psi$. Astfel, sînt posibile însumări ale produselor omogene din diferite întreprinderi însă sînt excluse însumările restului de produse din cadrul asortimentului aceleiași întreprinderi.

Elementul calitativ α nu se însumează niciodată. O asemenea însumare nu ar avea nici un sens economic, însă i se poate calcula valoarea medie:

$$\bar{\alpha} = \frac{\Sigma\alpha\psi}{\Sigma\psi} \quad \text{sau} \quad \bar{\alpha} = \frac{\Sigma'\omega}{\Sigma'\frac{\omega}{\alpha}}$$

cu condiția să existe $\Sigma'\psi$.

Categoriile economice pentru care elementul ψ nu se poate însuma vor fi denumite *obiecte* (producția eterogenă), iar cele pentru care acest element poate fi însumat vor fi denumite *părți omogene ale obiectului* sau, pe scurt, *părți ale obiectului* (producția omogenă în diferite întreprinderi).

Trebuie să atragem atenția că noțiunile de element funcțional ω sau element cantitativ ψ sînt considerate în sens relativ; indicatorii concreți pot juca rolul cînd uneia cînd alteia din mărimi, în funcție de condițiile concrete ale problemei.

Dacă considerăm, de exemplu, cheltuielile de timp de muncă T în activitatea unei întreprinderi, observăm că acest indicator apare ca element obligatoriu în relațiile din care se deduce indicele productivității muncii:

$$T = tq \qquad \text{(XV, 3.1)}$$

sau

$$q = wT, \qquad \text{(XV, 3.2)}$$

în care:

t este consumul de timp de muncă pe unitate de produs (consum specific);

q, cantitatea de produse de un anumit fel;

w, cantitatea de produse de un anumit fel, create într-o unitate de timp de muncă (productivitatea muncii).

În prima relație T apare ca element funcțional iar în a doua ca element cantitativ.

Același înțeles relativ au și noțiunile de *obiecte* și *părți ale obiectului*, precizate mai sus.

Folosind notația de mai sus pentru diferite categorii de indicatori statistici, vom examina funcțiile și construcția diferiților indici, delimitîndu-i — după V. Peregudov — în *patru grupe tipice.*

A. *Grupa indicilor care caracterizează cantitativ, dezvoltarea fenomenului complex în ansamblu.* (Indici ai indicatorului funcțional). Descompunerea mărimii de volum complexe ω în cadrul acestei grupe de indici, în elementele ei componente α și ψ, are numai un caracter potențial, în mișcarea lor, aceste elemente sînt considerate ca un tot. Cu alte cuvinte, în cadrul indicelui complex de tipul:

$$I_{\alpha,\psi} = \frac{\Sigma\omega_1}{\Sigma\omega_0} = \frac{\Sigma\alpha_1\psi_1}{\Sigma\alpha_0\psi_0} \qquad \text{(XV, 3.3)}$$

are loc un proces nemijlocit de trecere, în timp, de la $\alpha_0\psi_0$ spre $\alpha_1\psi_1$.

Exemplu de astfel de indice poate servi indicele circulației mărfurilor

$$I_{pq} = \frac{\Sigma p_1 q_1}{\Sigma p_0 q_0}.$$

B. *Grupa indicilor unui element izolat al fenomenului complex.* Indicii de acest tip caracterizează schimbarea în timp a unuia din elementele fenomenului complex, presupunînd că celălalt element rămîne invariabil, îndeplinind funcția de comensurare (ponderare). Elementul care nu variază și care servește drept pondere poate fi luat la nivelul lui din perioada de bază, din perioada curentă sau la oricare alt nivel. În funcție de aceasta pot fi considerate, deci, diferite variante de ponderare, pentru unul și același element component al fenomenului complex.

În practica actuală a statisticii s-a statornicit următorul principiu de ponderare: elementele *calitative* α se ponderează, de obicei, cu indicatorii *cantitativi* corespunzători din perioada curentă ψ_1 și invers — elementele de volum — *indicatorii cantitativi* ψ se ponderează cu *indicatorii calitativi* din perioada de bază α_0.

Formulele generale ale indicilor de acest tip:

$$I_{x(\psi_1)} = \frac{\Sigma x_1 \psi_1}{\Sigma x_0 \psi_1}; \quad I_{\psi(x_0)} = \frac{\Sigma x_0 \psi_1}{\Sigma x_0 \psi_0}; \quad I_{\psi(x)} = \frac{\Sigma x \psi_1}{\Sigma x \psi_0}, \quad \text{(XV, 3.4)}$$

își găsesc aplicații și în practica statisticii din țara noastră. În schimb, nu-și găsesc aplicații indicii de același tip (indici ai elementului izolat) la construirea cărora principiul ponderării formulat mai sus își schimbă sensul și anume:

$$I_{x(\psi_0)} = \frac{\Sigma x_1 \psi_0}{\Sigma x_0 \psi_0}; \quad I_{\psi(x_1)} = \frac{\Sigma x_1 \psi_1}{\Sigma x_1 \psi_0}. \quad \text{(XV, 3.5)}$$

Să aplicăm aceste formule generale la exemplul concret al elementelor componente ale volumului circulației mărfurilor.

Se știe că între volumul vînzărilor pe piață (v), prețurile (p) și cantitățile (q) ale fiecărui produs în parte există relația

$$v = pq.$$

Aceeași relație trebuie să existe și între indicii respectivi:

$$I_v = I_p \cdot I_q,$$

în cazul cînd se respectă principiul ponderare care duce la indici de tipul (XV, 3.4) sau (XV, 3.5).

Vom avea deci sistemul

$$I_v = \frac{\Sigma p_1 q_1}{\Sigma p_0 q_0} = \frac{\Sigma p_1 q_1}{\Sigma p_0 q_1} \cdot \frac{\Sigma p_0 q_1}{\Sigma p_0 q_0} \quad \text{(XV, 3.6)}$$

sau sistemul

$$I_v = \frac{\Sigma p_1 q_1}{\Sigma p_0 q_0} = \frac{\Sigma p_1 q_0}{\Sigma p_0 q_0} \cdot \frac{\Sigma p_1 q_1}{\Sigma p_1 q_0}. \quad \text{(XV, 3.7)}$$

Am arătat că în practică s-a adoptat sistemul (XV, 3.6) în care apare o mărime *convențională* $\Sigma q_1 p_0$ în combinarea elementelor calitativ (p) și cantitativ (q) ale fenomenului complex. Această mărime reprezintă volumul circulației mărfurilor din perioada curentă, exprimat în prețurile perioadei de bază. Sistemul (XV, 3.6) se deosebește de sistemul (XV, 3.7) prin faptul că în construcția indicilor pentru elementele constitutive din acest ultim sistem apare o altă mărime *convențională* $\Sigma p_1 q_0$.

După cum se vede, ambele sisteme sînt afectate de o anumită convenționalitate; în construcția lor apar anumite mărimi *nereale*, ceea ce ar constitui un motiv puternic pentru a le respinge, pe amîndouă. Analiza teoretică ne duce, însă, la o altă concluzie, aceea a necesității de a fi aplicată metoda izolării unuia din factori pentru a studia schimbarea în timp a celuilalt. Se pune atunci întrebarea, căreia din cele două mărimi convenționale trebuie să i se dea preferință și, în consecință, care din cele două sisteme trebuie adoptat în practică? Întrucît din punct de vedere

formal nu avem nici un motiv de a da preferință unuia din cele două sisteme examinate, răspunsul corect la întrebarea pusă ni-l poate da numai *analiza social-economică concretă* a fiecărui fenomen în parte.

În procesul unei asemenea analize sîntem adeseori nevoiți a calcula întregul sistem de indici și totuși, în fiecare caz concret de aplicare a metodei indicilor, *una din verigile sistemului ne interesează îndeosebi* deși, în general, se calculează întregul sistem de indici; în unele cazuri ne va interesa dinamica circulației mărfurilor sub influența schimbării prețurilor, în altele, aceeași dinamică, în funcție de schimbarea volumului de mărfuri. Aceasta ar fi prima împrejurare de care trebuie să ținem seama cînd sîntem puși în situația de a alege unul din cele două sisteme de calcul. Pe de altă parte, nu ne este indiferent nici faptul dacă ne propunem drept scop de a oglindi dinamica fenomenului sub aspectul dezvoltării lui în momentul actual sau acela de a caracteriza mai pregnant o anumită perioadă trecută.

Să presupunem că ne stă în față sarcina de a analiza indicatorul dinamicii relative a circulației mărfurilor pentru momentul actual *relevînd influența pe care o exercită dinamica prețurilor*. Este evident că în acest caz, indicatorul principal asupra căruia ne vom îndrepta atenția este acela pe care-l denumim, în mod obișnuit, indicele prețurilor, iar volumul circulației mărfurilor din perioada actuală va fi în mod obligatoriu o mărime *reală* — adică $\Sigma p_1 q_1$ (la determinarea ei nu vom admite nici un fel de convenționalitate). Devine clar că folosirea în acest caz a sistemului de indici (XV, 3.6) se impune cu necesitate. Tocmai de aceea, în practica actuală a statisticii socialiste, ori de cîte ori este vorba de a fi calculat indicele prețurilor, este folosit primul sistem al indicilor (XV, 3.6). Necesitățile practice au dus, de altfel, la formularea principiului de ponderare statornicit în practica statisticii socialiste, arătat mai sus.

În aceleași condiții, dacă ne interesează, îndeosebi, influența pe care o exercită dinamica volumului de mărfuri (este vorba tot de influența asupra dinamicii relative a indicatorului circulației mărfurilor *din perioada actuală*), este evident că, în acest caz, vom folosi al doilea sistem al indicilor (XV, 3.7).

Dacă, dimpotrivă, în față ne stă sarcina de a compara cu situația actuală, condițiile circulației mărfurilor dintr-o perioadă trecută, să zicem din perioada anului 1907 (este vorba, deci, de o cercetare cu caracter istoric, care are drept scop de a prezenta o imagine cît mai sugestivă pentru cercetătorul contemporan), este clar că în acest caz nu ne-ar fi îngăduit de a opera cu anumiți indicatori convenționali în ceea ce privește circulația mărfurilor din perioada trecută. Va fi folosit deci sistemul în care apare indicatorul real $\Sigma p_0 q_0$. De aceea, în funcție de locul pe care-l va ocupa în optica cercetării noastre unul din elementele circulației mărfurilor, vom acorda prioritate unuia sau altuia din cele două sisteme examinate.

Din exemplul concret al indicilor circulației mărfurilor, prețurilor și al volumului de mărfuri, putem trage următoarele concluzii:

1°. Alegerea *sistemului* de indici trebuie să fie determinată de analiza științifică a problemei de rezolvat, în funcție de condițiile ei concrete și

de sarcinile cognitive pe care trebuie să le îndeplinească indicatorii relativi ai dinamicii.

2°. Este vorba, întotdeauna, de alegerea unui *sistem* de indici și nu de alegerea unei formule de calcul pentru fiecare caracteristică în parte.

3°. Numărul variantelor de sisteme de calcul al indicilor crește mereu, dacă fenomenul complex se descompune în elemente tot mai multe, devenind de forma generală $\omega = \alpha\beta\gamma \ldots$ Vom avea astfel, în cazul unui fenomen complex format din trei elemente, un număr de șase sisteme diferite de indici, în cazul a patru elemente, — 24 sisteme de indici iar în cazul a n elemente, un număr de $n!$ sisteme de indici *).

Mai socotim util a arăta că pentru indicii elementelor constitutive ale unui fenomen complex, de tipul B (pe care-l examinăm) sînt folosite în practică și formule transformate în care se consideră indicii individuali $i_\alpha = \frac{\alpha_i}{\alpha_0}$ și $i_\psi = \frac{\psi_i}{\psi_0}$.

Drept ponderi pentru acești indici individuali sînt luate mărimile efective ale indicatorilor funcționali: $\omega_0 = \alpha_0\psi_0$ sau $\omega_1 = \alpha_1\psi_1$, folosiți la construirea indicilor de tipul B care se mai numesc și *indici agregați* (în sens general, noțiunea de „agregat" se referă la suma de mărimi cu structuri diferite; prin denumirea de „indice agregat" se scoate în evidență acea particularitate a unui indice de a fi indicatorul dinamicii unor fenomene de masă cu structuri diferite). Calculul se efectuează pe baza formulei mediei aritmetice sau mediei armonice:

$$I_{\alpha(\psi)} = \frac{\Sigma\alpha_1\psi_1}{\Sigma\alpha_0\psi_1} = \frac{\Sigma\omega_1}{\Sigma\omega_1/i_\alpha} = \frac{\Sigma\alpha_1\psi_1}{\sum\frac{1}{i_\alpha}\alpha_1\psi_1} \qquad \text{(XV, 3.8)}$$

sau

$$I_{\psi(\alpha)} = \frac{\Sigma\alpha_0\psi_1}{\Sigma\alpha_0\psi_0} = \frac{\Sigma i_\psi\omega_0}{\Sigma\omega_0}. \qquad \text{(XV, 3.9)}$$

De exemplu, revenind la exemplul concret al indicilor elementelor constitutive ale circulației mărfurilor, avem

$$I_{p(q_1)} = \frac{\Sigma p_1q_1}{\Sigma p_0q_1} = \frac{\Sigma p_1q_1}{\Sigma p_1q_1/i_p} = \frac{\Sigma p_1q_1}{\sum\frac{1}{i_p}p_1q_1}. \qquad \text{(XV, 3.10)}$$

Este ușor de observat că această transformare algebrică a formulei indicelui prețurilor din sistemul (XV, 3.6) ne permite să tragem concluzia că indicele agregat al prețurilor se calculează pe baza mediei armonice, ori de cîte ori în practică ni se oferă numai date asupra indicilor individuali ai prețurilor și asupra volumului circulației mărfurilor din perioada curentă.

*) Vezi în acest sens și L. V. Nekraș, *Kurs obșcei teorii statistiki*. Gosplanizdat, Moscova-Leningrad, 1939, p. 283—284.

Forma mediei aritmetice va fi folosită dacă avem date asupra indicilor individuali ai prețurilor și asupra volumului convențional al circulației mărfurilor din perioada curentă ($\Sigma p_0 q_1$) :

$$I_{p(q_1)} = \frac{\Sigma p_1 q_1}{\Sigma p_0 q_1} = \frac{\sum \frac{p_1}{p_0} p_0 q_1}{\Sigma p_0 q_1} = \frac{\Sigma i_p p_0 q_1}{\Sigma p_0 q_1}. \qquad \text{(XV, 3.11)}$$

Pentru indicele volumului de mărfuri (q) avem formula

$$I_{q(p_0)} = \frac{\Sigma p_0 q_1}{\Sigma p_0 q_0} = \frac{\sum \frac{q_1}{q_0} p_0 q_0}{\Sigma p_0 q_0} = \frac{\Sigma i_q p_0 q_0}{\Sigma p_0 q_0}. \qquad \text{(XV, 3.12)}$$

Scrierea indicilor sub forma unor medii aritmetice sau armonice prezintă și o altă utilitate. Cu ajutorul lor putem să ne explicăm unele rezultate. Astfel, aplicînd formula (XV, 3.11), vom constata că valoarea indicelui $I_{p(q_1)}$ este cel mai mult influențată de indicii i_p pentru care ponderea corespunzătoare $p_0 q_1$ este cea mai mare.

Pentru cel de al doilea sistem al indicilor (XV, 3.7), de care ne vom folosi dacă în optica dominantă a cercetării noastre se află schimbarea volumului și a structurii volumului de mărfuri, indicii agregați respectivi vor lua următoarea formă :

$$I_{p(q_0)} = \frac{\Sigma p_1 q_0}{\Sigma p_0 q_0} = \frac{\Sigma i_p p_0 q_0}{\Sigma p_0 q_0} \qquad \text{(XV, 3.13)}$$

și

$$I_{q(p_1)} = \frac{\Sigma p_1 q_1}{\Sigma p_1 q_0} = \frac{\Sigma p_1 q_1}{\sum \frac{1}{i_q} p_1 q_1}. \qquad \text{(XV, 3.14)}$$

Să considerăm un exemplu de calcul al sistemului de indici ai circulației mărfurilor (tabelul 62).

Tabelul 62

Produse	U/m	Perioada de bază: Prețuri pe unitate de produs p_0	Perioada de bază: Cantități în unități de măsură arătate q_0	Perioada de bază: Circulația mărfurilor în lei p_0q_0	Perioada curentă: Prețuri pe unitate de produs p_1	Perioada curentă: Cantități în unități de măsură arătate q_1	Perioada curentă: Circulația mărfurilor în lei p_1q_1	Circulația mărfurilor convențională (în prețuri per bază) p_0q_1	Circulația mărfurilor convențională (preț. curente și cantități din bază. p_1q_0	Indicii individuali ai prețurilor p_1/P_0	Mărimi inverse ale indicilor individuali p_0/q_1
A	B	1	2	3	4	5	6	7	8	9	10
A	kg	180	250	45 000	160	400	64 000	72 000	40 000	0,889	1,125
B	m	30	5 000	150 000	25	6 000	150 000	180 000	125 000	0,833	1,200
C	buc.	70	2 500	175 000	80	1 500	120 000	105 000	200 000	1,143	0,875
Total	—	—	—	370 000	—	—	334 000	357 000	365 000	—	—

În acest tabel avem toate datele necesare pentru a calcula indicatorii de care avem nevoie:

Să calculăm mai întîi sistemul de indici (XV, 3.6)

$$I_v = \frac{\Sigma p_1 q_1}{\Sigma p_0 q_0} = \frac{334\,000}{370\,000} = 0{,}903 \quad \text{sau} \quad 90{,}3\%;$$

$$I_q(q_1) = \frac{\Sigma p_1 q_1}{\Sigma p_0 q_1} = \frac{334\,000}{357\,000} = 0{,}935 \quad \text{sau} \quad 93{,}5\%;$$

$$I_p(p_0) = \frac{\Sigma p_0 q_1}{\Sigma p_0 q_0} = \frac{357\,000}{370\,000} = 0{,}965 \quad \text{sau} \quad 96{,}5\%.$$

Rezultatul calculelor ne arată că volumul total al circulației mărfurilor în perioada curentă a scăzut cu 9,7%. Această scădere a avut loc sub influența a doi factori:

a) la produsele *A* și *B* au scăzut prețurile (vezi coloana 9 în care valorile indicilor individuali ai prețurilor la aceste produse a scăzut sub 1) și totodată a crescut volumul produselor desfăcute;

b) a crescut prețul la produsul *C* însă a scăzut considerabil volumul cantităților desfăcute.

Rezultă, în ansamblu, că sub influența scăderii prețurilor volumul circulației mărfurilor a scăzut cu 6,5%, față de volumul său din perioada de bază și apoi, sub influența schimbării cantității totale de mărfuri desfăcute, acest volum a mai scăzut cu 3,5%.

Controlul calculelor ne arată exactitatea lor

$$0{,}935 \cdot 0{,}965 \cong 0{,}902 \text{ sau } 90{,}2\%.$$

Folosind elementele indicelui prețurilor mai putem obține volumul absolut al schimbării valorii produselor desfăcute pe seama schimbării prețurilor. Se vede că el este dat de diferența dintre numărător și numitor egală cu $334\,000 - 357\,000 = -23\,000$ lei și reprezintă economiile realizate de cumpărători prin efectul reducerii prețurilor în perioada curentă.

Să calculăm acum indicii din sistemul (XV, 3.7):

$$I_v = \frac{\Sigma p_1 q_1}{\Sigma p_0 q_0} = \frac{334\,000}{370\,000} = 0{,}903 = 90{,}3\%,$$

$$I_p(q_0) = \frac{\Sigma p_1 q_0}{\Sigma p_0 q_0} = \frac{365\,000}{370\,000} = 0{,}986 = 98{,}6\%,$$

$$I_q(p_1) = \frac{\Sigma p_1 q_1}{\Sigma p_1 q_0} = \frac{334\,000}{365\,000} = 0{,}915 = 91{,}5\%.$$

Se observă că dacă în primul caz, datorită schimbării prețurilor, volumul circulației mărfurilor a scăzut cu 6,5% iar datorită schimbării cantităților de mărfuri desfăcute, cu 3,5% în acest caz, el a scăzut respectiv cu 1,4% și 8,5%.

Sistemul se verifică însă şi în acest caz :

$$0{,}986 \cdot 0{,}915 = 0{,}9022 \cong 0{,}902 \text{ sau } 90{,}2\%.$$

Diferența de valori înregistrată la indicii elementelor componente ale indicatorului complex nu poate fi explicată decît prin aceea că, la construirea *indicatorilor dinamicii convenționale* ($\Sigma p_0 q_1$ *şi* $\Sigma p_1 q_0$) *noi ignorăm legătura reciprocă reală, care există între elementele* p_i *şi* q_i *şi operăm cu ele ca şi cum ar fi independente.*

Se pune deci problema de a fi găsite soluțiile corespunzătoare, în primul rînd pe plan teoretic, pentru a fi în general evitate unele contradicții ce se observă la aplicarea metodei indicilor.

Vom arăta mai jos unele soluții propuse de V. Peregudov în această problemă.

Am arătat mai sus cîteva considerente care au determinat acceptarea în practică a indicelui prețurilor din sistemul (XV, 3.6). Mai putem invoca acum şi alte considerente :

În primul rînd, se pune o problemă de interpretare a însuşi sensului acestui indice. Denumindu-l „indice al prețurilor" trebuie să avem totuşi în vedere faptul incontestabil că el nu reprezintă, în esență, un indicator al schimbării „nivelului" prețurilor, ci tot *un indicator al dinamicii relative a circulației mărfurilor sub influența schimbării prețurilor* şi de aceea în condițiile în care celălalt element component (q) rămîne constant, trebuie să raționăm în felul următor : Elementul prim — punctul de plecare — al comparării în timp trebuie să-l constituie volumul circulației mărfurilor din *perioada curentă*, în condițiile schimbării prețurilor şi a structurii volumului de mărfuri format în această perioadă ($\Sigma p_1 q_1$), întrucît procedînd altfel — construind pentru perioada curentă indicatorul convențional al circulației mărfurilor $\Sigma p_1 q_0$ şi comparîndu-l apoi cu indicatorul real din perioada de bază $\Sigma p_0 q_0$ (aşa după cum se procedează la calculul indicelui prețurilor din sistemul (XV, 3.7)) — nu vom ține seama de schimbările care intervin în structura circulației mărfurilor pe parcurs de la perioada de bază la perioada curentă. Or, tocmai aceasta ar duce la interpretări greşite asupra dinamicii relative a volumului valoric al circulației mărfurilor, sub influența schimbării prețurilor.

De aceea, practic, sîntem obligați să alegem acel sistem de indici care include „indicele prețurilor" de forma :

$$I_p(q_1) = \frac{\Sigma p_1 q_1}{\Sigma p_0 q_1}.$$

Dimpotrivă, dacă sarcinile problemei ce o supunem studiului ne-ar impune să acordăm atenție dominantă schimbărilor care intervin în volumul mărfurilor, în acest caz ar trebui să judecăm în mod analog în ceea ce priveşte aşa-numitul „indice al volumului fizic al mărfurilor" şi să ne pronunțăm, în consecință, pentru sistemul de indici (XV, 3.6) în care el apare sub forma :

$$I_q(p_0) = \frac{\Sigma p_0 q_1}{\Sigma p_0 q_0}.$$

Scanned with OKEN Scanner

Este evident că pentru a alege un sistem sau altul, pe primul plan în acest scop ne apar considerațiile cu privire la indicele prețurilor, legate nemijlocit de nevoile noastre concrete, practice, deoarece volumul de mărfuri nefiind măsurabil direct, nici un fel de aprecieri asupra volumului ei nu se pot face înainte de a-l exprima în unități bănești.

Referindu-ne la tabelul 62 și presupunînd că ne-ar sta la dispoziție numai datele din coloanele 3, 6, 7, 8, 9 și 10, vom putea calcula cu ușurință indicii de forma mediei aritmetice sau armonice, folosindu-ne de formulele (XV, 3.10), (XV, 3.11) și (XV, 3.12).

De exemplu, indicele prețurilor calculat pe baza formulei mediei armonice va fi egal cu :

$$I_p = \frac{\Sigma p_1 q_1}{\sum \frac{1}{i_p} p_1 q_1} = \frac{334\,000}{64\,000 \cdot 1{,}125 + 150\,000 \cdot 1{,}200 + 120\,000 \cdot 0{,}875} =$$

$$= 0{,}935 \text{ sau } 93{,}5\%.$$

Rezultatul obținut este identic cu cel dat de formula indicelui agregat din sistemul (XV, 1.6).

Același rezultat se obține aplicînd formula mediei aritmetice (XV, 3.11) :

$$I_p = \frac{\Sigma i_p p_0 q_1}{\Sigma p_0 q_1} = \frac{0{,}889 \cdot 72\,000 + 0{,}833 \cdot 180\,000 + 1{,}143 \cdot 150\,000}{72\,000 + 180\,000 + 105\,000} =$$

$$= 0{,}935 \text{ sau } 93{,}5\%.$$

Nu este greu de controlat că și la indicii volumului de mărfuri (I_q), calculați cu ajutorul formulelor transformate, obținem rezultate identice cu cele ce ni le dau formulele indicilor agregați.

În legătură cu aplicarea indicilor agregați se ridică o problemă importantă și anume *problema comparării datelor pe mai multe perioade.*

Să examinăm modul în care se rezolvă această problemă, în cazul exemplului cu circulația mărfurilor.

Fie

$$p_1, p_2, \ldots, p_n$$

prețurile unui anumit produs în n perioade și

$$q_1, q_2, \ldots, q_n$$

cantitățile aceluiași produs în n perioade, și că avem o întreagă colectivitate de produse pentru care vom avea, evident, prețurile și cantitățile respective tot pe n perioade. Pentru totalitatea volumului circulației mărfurilor, în fiecare perioadă, vom avea mărimile : $\Sigma p_1 q_1$; $\Sigma p_2 q_2$; ... ; $\Sigma p_n q_n$.

Indicii respectivi vor putea fi construiți în două feluri :

1°. Se aleg datele unei perioade drept bază de comparație și apoi se compară cu ele datele fiecărei perioade succesive; avem deci o bază constantă în calculul tuturor indicilor. Indicii construiți prin această metodă sînt, în fond, indici agregați în formă de bază.

2°. Baza de comparație se schimbă de la un indice la altul și datele folosite în acest scop sînt întotdeauna datele perioadei precedente celei pentru care se construiește indicele. După cum se vede, indicii obținuți în acest mod sînt indici agregați în formă de lanț.

Astfel, indicii agregați cu bază constantă au pentru indicatorul funcțional al circulației mărfurilor următoarea formă:

$$\frac{\Sigma p_2 q_2}{\Sigma p_1 q_1}, \frac{\Sigma p_3 q_3}{\Sigma p_1 q_1}, \ldots, \frac{\Sigma p_n q_n}{\Sigma p_1 q_1} \qquad \text{(XV, 3.15)}$$

iar indicii în formă de lanț:

$$\frac{\Sigma p_2 q_2}{\Sigma p_1 q_1}, \frac{\Sigma p_3 q_3}{\Sigma p_2 q_2}, \ldots, \frac{\Sigma p_n q_n}{\Sigma p_{n-1} q_{n-1}}. \qquad \text{(XV, 3.16)}$$

O altă formă capătă indicii agregați ai prețurilor și ai cantităților de mărfuri — elementele componente ale sistemului (XV, 3.6):

a) indicii prețurilor în formă de bază

$$\frac{\Sigma p_2 q_2}{\Sigma p_1 q_2}, \frac{\Sigma p_3 q_3}{\Sigma p_1 q_3}, \frac{\Sigma p_4 q_4}{\Sigma p_1 q_4}, \ldots, \frac{\Sigma p_n q_n}{\Sigma p_1 q_n}; \qquad \text{(XV, 3.17)}$$

b) indicii prețurilor în formă de lanț

$$\frac{\Sigma p_2 q_2}{\Sigma p_1 q_2}, \frac{\Sigma p_3 q_3}{\Sigma p_2 q_3}, \frac{\Sigma p_4 q_4}{\Sigma p_3 q_4}, \ldots, \frac{\Sigma p_n q_n}{\Sigma p_{n-1} q_n}; \qquad \text{(XV, 3.18)}$$

a) indicii cantităților în formă de bază

$$\frac{\Sigma p_1 q_2}{\Sigma p_1 q_1}, \frac{\Sigma p_1 q_3}{\Sigma p_1 q_1}, \frac{\Sigma p_1 q_4}{\Sigma p_1 q_1}, \ldots, \frac{\Sigma p_1 q_n}{\Sigma p_1 q_1}; \qquad \text{(XV, 3.19)}$$

b) indicii cantităților în formă de lanț

$$\frac{\Sigma p_1 q_2}{\Sigma p_1 q_1}, \frac{\Sigma p_2 q_3}{\Sigma p_2 q_2}, \frac{\Sigma p_3 q_4}{\Sigma p_3 q_3}, \ldots, \frac{\Sigma p_{n-1} q_n}{\Sigma p_{n-1} q_{-1}}. \qquad \text{(XV, 3.20)}$$

Se poate ușor controla că într-adevăr indicii construiți se leagă în sistemul de tip (XV,3.6). Să luăm, de exemplu, indicii în formă de lanț, pentru anul al patrulea și vom avea

$$Iv, 4/3 = Ip, 4/3 \cdot Iq, 4/3 = \frac{\Sigma p_4 q_4}{\Sigma p_3 q_3} = \frac{\Sigma p_4 q_4}{\Sigma p_3 q_4} \cdot \frac{\Sigma p_3 q_4}{\Sigma p_3 q_3},$$

sau indicii în formă de bază

$$Iv, 4/1 = Ip, 4/1 \cdot Iq, 4/1 = \frac{\Sigma p_4 q_4}{\Sigma p_1 q_1} = \frac{\Sigma p_4 q_4}{\Sigma p_1 q_4} \cdot \frac{\Sigma p_1 q_4}{\Sigma p_1 q_1}.$$

Observăm că indicii în formă de lanț pentru indicatorul funcțional (XV, 3.16) se bucură de proprietatea de trecere la indicii de bază și că de această

proprietate nu se bucură indicii în formă de lanț ai prețurilor (XV, 3.17) și cantităților (XV, 3.20).

Într-adevăr,

$$\frac{\Sigma p_2 q_2}{\Sigma p_1 q_1} \cdot \frac{\Sigma p_3 q_3}{\Sigma p_2 q_2} \cdot \frac{\Sigma p_4 q_4}{\Sigma p_3 q_3} \cdots \frac{\Sigma p_{n-1} q_{n-1}}{\Sigma p_{n-2} q_{n-2}} \cdot \frac{\Sigma p_n q_n}{\Sigma p_{n-1} q_{n-1}} = \frac{\Sigma p_n q_n}{\Sigma p_1 q_1}, \quad \text{(XV, 3.21)}$$

$$\frac{\Sigma p_2 q_2}{\Sigma p_1 q_2} \cdot \frac{\Sigma p_3 q_3}{\Sigma p_2 q_3} \cdots \frac{\Sigma p_n q_n}{\Sigma p_{n-1} q_n} \neq \frac{\Sigma p_1 q_n}{\Sigma p_1 q_1}, \quad \text{(XV, 3.22)}$$

$$\frac{\Sigma p_1 q_2}{\Sigma p_1 q_1} \cdot \frac{\Sigma p_2 q_3}{\Sigma p_2 q_2} \cdots \frac{\Sigma p_{n-1} q_n}{\Sigma p_{n-1} q_{n-1}} \neq \frac{\Sigma p_1 q_n}{\Sigma p_1 q_1}. \quad \text{(XV, 3.23)}$$

Pentru ca inegalitățile (XV, 3.22) și (XV, 3.23) să se poată transforma în egalități este suficient ca ponderile folosite să fie *constante* și anume, cele din prima sau ultima perioadă. De exemplu, folosind ponderi constante, inegalitatea (XV, 3.22) se transformă în egalitate:

$$\frac{\Sigma p_2 q_1}{\Sigma p_1 q_1} \cdot \frac{\Sigma p_3 q_1}{\Sigma p_2 q_1} \cdot \frac{\Sigma p_4 q_1}{\Sigma p_3 q_1} \cdots \frac{\Sigma p_n q_1}{\Sigma p_{n-1} q_1} = \frac{\Sigma p_n q_1}{\Sigma p_1 q_1}. \quad \text{(XV. 3.24)}$$

Dar această condiție este departe de a fi și necesară.

Bunăoară, dacă luăm $n = 3$, din (XV, 3.22) deducem

$$\frac{\Sigma p_2 q_2}{\Sigma p_1 q_2} = \frac{\Sigma p_2 q_3}{\Sigma p_1 q_3}.$$

Se vede că sistemele de ponderi q_2 și q_3 trebuie să fie astfel încît să conducă la același indice al prețurilor. Aceasta reprezintă însă o structură particulară a volumului mărfurilor desfăcute, care nu poate fi luată în considerare la alcătuirea unei teorii generale.

Dacă construcția indicilor cu ponderi constante, ca cei din relația (XV, 3.24), satisface proprietatea de trecere a indicilor agregați de la forma lor în lanț la forma de bază, nu este mai puțin adevărat că o asemenea construcție nu are nici un sens economic. Într-adevăr, ne putem întreba, ce sens economic ar avea calculul volumului valoric al circulației mărfurilor în fiecare an, considerînd cantitățile de mărfuri desfăcute în anul inițial (de bază) al seriei studiate? Aceasta ar însemna să presupunem — în pofida dezvoltării reale a fenomenului — că structura și volumul mărfurilor desfăcute, an de an, pe o perioadă de timp relativ lungă rămîn aceleași. Or, este evident că o asemenea presupunere nu ar avea nici un conținut real, nici un sens economic.

Cu totul altă imagine ne oferă indicii în formă de lanț construiți pe baza ponderii constante a ultimei perioade:

$$\frac{\Sigma p_2 q_n}{\Sigma p_1 q_n} \cdot \frac{\Sigma p_3 q_n}{\Sigma p_2 q_n} \cdots \frac{\Sigma p_{n-1} q_n}{\Sigma p_{n-2} q_n} \cdot \frac{\Sigma p_n q_n}{\Sigma p_{n-1} q_n} = \frac{\Sigma p_n q_n}{\Sigma p_1 q_n} \quad \text{(XV, 3.25)}$$

După cum se vede, de data aceasta am obținut un indice de bază care satisface condițiile de ponderare adoptate în general la construirea indi-

cilor dinamicii relative a circulației mărfurilor sub influența schimbării prețurilor.

Trebuie să menționăm că indicii agregați în formă de lanț calculați pentru indicatorul funcțional se bucură întotdeauna de proprietatea (XV, 3.21) dacă funcția determinantă este de tipul $\omega = \alpha\,\psi$.

C. *Grupa indicilor care caracterizează dinamica medie a unui singur element* (*indicii cu structură variabilă*). Din această grupă fac parte indicii cu ajutorul cărora se realizează compararea în timp a două mărimi medii, pentru al căror calcul se poate însuma elementul de volum — indicatorul cantitativ ψ.

Formulele generale de calcul al acestor indici sînt de tipul :

$$I_{\bar{\alpha}} = \frac{\bar{\alpha}_1}{\bar{\alpha}_0} \text{ pentru indicatorul calitativ și}$$

$$I_{\psi} = \frac{\Sigma' \psi_1}{\Sigma' \psi_0} \text{ pentru indicatorul cantitativ.} \qquad \text{(XV, 3.26)}$$

Indicile de forma (XV, 3.26) este cunoscut în statistica socialistă sub denumirea de *indice cu structură variabilă.*

Adeseori, pe baza mediilor $\bar{\alpha}$ și a sumelor $\Sigma' \psi = \psi$ se calculează un indice mai complex, potrivit schemei indicilor de tipul B :

$$I_{\bar{\alpha}(\psi)} = \frac{\Sigma \bar{\alpha}_1 \Psi_1}{\Sigma \bar{\alpha}_0 \Psi_1}. \qquad \text{(XV, 3.27)}$$

Indicele de tipul (XV, 3.27) se folosește la calculul indicelui prețurilor, pentru mai multe mărfuri și, totodată, pe mai multe orașe, regiuni etc. De asemenea, indici de acest tip vom întîlni și la calculul indicilor teritoriali.

Vom ilustra modul de calcul al acestor indici cu prilejul expunerii asupra indicilor prețului de cost și ai productivității muncii, cînd ne vom întîlni și cu indici de tipul celor din grupele A, B și D. Analizei indicilor cu structură variabilă îi vom consacra un paragraf aparte.

D. *Grupa indicilor care caracterizează deplasările structurale.* La construirea indicilor de tipul D se compară structurile diferite în timp ale unei colectivități de obiecte, cu ajutorul unităților de comensurare (ponderare) reprezentate prin indicatorii calitativi α

Indicatorul calitativ α poate fi luat la nivelul lui din perioada de bază (α_0) sau la cel din perioada curentă, (α_1).

Formula generală a indicilor de tipul grupei D va fi deci

$$I_{\psi^{*}(\alpha_0)} = \frac{\Sigma' \alpha_0 \Psi^{*}_1}{\Sigma' \alpha_0 \Psi^{*}_0} \qquad \text{(XV, 3.28)}$$

sau

$$I_{\psi^{*}(\alpha_1)} = \frac{\Sigma' \alpha_1 \Psi^{*}_1}{\Sigma' \alpha_1 \Psi^{*}_0} \qquad \text{(XV, 3.29)}$$

în care

$$\Psi_1^* = \frac{\Psi_1}{\Sigma'\Psi_1}, \quad \Psi_0^* = \frac{\Psi_0}{\Sigma'\Psi_0}.$$

În practică s-a adoptat indicele deplasărilor structurale de forma (XV, 3.29).

Un exemplu concret ni-l poate oferi indicele care caracterizează factorul teritorial și schimbarea prețurilor:

$$I_{q^*(p_0)} = \frac{\Sigma' p_0 q_1^*}{\Sigma' p_0 q_0^*}.$$

Un medel mai complex al indicelui de acest tip ar fi:

$$I_{\psi^*(\alpha_0)} = \frac{\Sigma(\Psi_1 \cdot \Sigma'\alpha_0\Psi_1^*)}{\Sigma(\Psi_1 \cdot \Sigma'\alpha_0\Psi_0^*)}. \qquad \text{(XV, 3.30)}$$

Aplicațiilor practice ale indicilor de acest tip le vom consacra un paragraf separat în cadrul analizei indicelui cu structură variabilă.

*

Trebuie să remarcăm că gruparea indicilor în cele patru tipuri nu are un caracter rigid. În unele condiții este posibilă o deplină identitate a indicilor care din punct de vedere formal fac parte din grupe diferite. De exemplu, dacă unul din elementele fenomenului complex rămîne neschimbat de la o perioadă la alta (să zicem $\alpha_1 = \alpha_0$), în acest caz indicele de tipul A devine identic cu indicele corespunzător din grupa B. Sau, indicele cu structură variabilă (de tipul indicilor din grupa C) poate fi considerat ca un caz particular al indicilor de tipul A întrucît:

$$\bar{\alpha}_1 = \frac{\Sigma'\alpha_1\psi_1}{\Sigma'\psi_1} = \Sigma'\alpha_1\psi_1^* \quad \text{și} \quad \bar{\alpha}_0 = \frac{\Sigma'\alpha_0\psi_0}{\Sigma'\psi_0} = \Sigma'\alpha_0\psi_0^*.$$

Se mai pot da și alte exemple de identitate a indicilor aparținînd grupelor diferite. Aceasta nu reduce însă cu nimic utilitatea practică a acestei clasificații, propusă de V. Peregudov, dat fiind că ea se referă tocmai la sistemele de indici care se aplică de obicei în practica statisticii socialiste.

4. Aplicații ale calculului indicilor. Indici ai prețului de cost

Ne vom opri asupra unor particularități pe care le întîlnim în calculul indicilor prețului de cost în industrie, ținînd seama de tipurile de indici examinați în paragraful precedent și de principiile teoretice adoptate în practica lucrărilor statistice.

Scanned with OKEN Scanner

Tabelul 63

Produse	Preț de cost pe unitate de produs (lei)			Numărul produselor (în unități de măsură naturale)			Costul producției marfă comparabilă (planificată, efectivă și convențională) (în lei)						Indicii individuali ai prețurilor
	z_0	z_{pl}	z_1	q_0	q_{pl}	q_1	$z_0 q_{pl}$	$z_{pl} q_{pl}$	$z_1 q_{pl}$	$z_0 q_1$	$z_{pl} q_1$	$z_1 q_0$	$\frac{z_1}{z_0} = i$
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
A	100	90	88	500	600	1 000	60 000	54 000	52 800	100 000	90 000	88 000	50 000	44 000	0,88
B	50	44	48	400	240	200	12 000	10 560	11 520	10 000	8 800	9 600	20 000	19 200	0,96
C	30	25	20	300	500	800	15 000	12 500	10 000	24 000	20 000	16 000	9 000	6 000	0,66
Total	—	—	—	—	—	—	87 000	77 060	74 320	134 000	118 800	113 600	79 000	69 200	—

Vom considera un tabel (tabelul 63) cu date (convenționale) ale unei intreprinderi industriale, caracterizînd elementele ce ne sînt necesare pentru calculul diferiților indici ai prețului de cost.

Pe baza datelor din tabelul 63, putem calcula, în primul rînd, *indicele sarcinii de reducere a prețului de cost*, în care datele de plan se asimilează „perioadei curente". Fiind vorba de indicele unui element calitativ din relația determinantă $r = zq$ (în care r reprezintă cheltuielile de producție, z, prețul de cost unitar și q numărul fiecărui fel de produs), se va aplica formula indicelui agregat de tipul B, (XV, 3.4):

$$I_{zpl(qpl)} = \frac{\Sigma z_{pl}\, q_{pl}}{\Sigma z_0\, q_{pl}} = \frac{77\,060}{87\,000} =$$

$$= 0{,}8857 \text{ sau } 88{,}57\%. \quad (XV, 4.1)$$

Se vede că sarcina planificată de reducere a prețului de cost este de 11,43%, iar economiile planificate se obțin prin diferența între numărător și numitor;

$$\Sigma_{zplqpl} - \Sigma z_0\, q_{pl} =$$

$$= 770\,060 - 87\,000 = -\,9\,940.$$

În al doilea rînd, pe baza acelorași date se calculează ușor indicele dinamicii relative a prețului de cost al producției față de nivelul lui din perioada precedentă, cu folosirea ponderilor de plan

$$I_{z(qpl)} = \frac{\Sigma z_1\, q_{pl}}{\Sigma z_0 q_{pl}} = \frac{74\,320}{87\,000} =$$

$$= 0{,}854 \text{ sau } 85{,}4\%. \quad (XV, 4.2)$$

Acest indice ne arată că dacă întreprinderea ar fi respectat structura de plan a producției la nivelul dat al prețului de cost (z_1) al diferitelor produse, atunci reducerea medie a prețului de cost a producției marfă

comparabilă sub influența reducerii prețului de cost a diferitelor produse ar fi de 14,60%, adică mai mare decît sarcina de plan (vezi (XV, 4.1)).

După cum se vede, un asemenea indice are o funcție analitică deosebit de importantă.

În al treilea rînd se poate calcula gradul de îndeplinire a planului prețului de cost, în două variante:

a) se compară prețul de cost efectiv cu cel planificat folosindu-se ponderile planului (q_{pl}):

$$I_{z(qpl)} = \frac{\Sigma z_1 q_{pl}}{\Sigma z_{pl} q_{pl}} = \frac{74\,320}{77\,060} = 0{,}9644 \quad \text{sau} \quad 96{,}44\%. \quad (XV, 4.2\,a)$$

Acest indice ne spune că, sub influența reducerii prețului de cost al celor trei produse, prețul de cost al producției marfă comparabile (cheltuielile de producție) a scăzut cu 3,56%, realizîndu-se un volum de economii peste plan egal cu $\Sigma z_1 q_{pl} - \Sigma z_{pl} q_{pl} = 74\,320 - 77\,060 = -2\,740$.

b) Se compară indicele prețului de cost efectiv în construcția căruia au intrat ponderile de plan, cu indicele sarcinii de plan (XV, 2.1), adică se construiește raportul:

$$I_{z(qpl)} = \frac{\Sigma z_1 q_{pl}}{\Sigma z_0 q_{pl}} : \frac{\Sigma z_{pl} q_{pl}}{\Sigma z_0 q_{pl}} = \frac{\Sigma z_1 q_{pl}}{\Sigma z_{pl} q_{pl}} = \frac{74\,320}{87\,000} : \frac{77\,060}{87\,000} =$$

$$= 0{,}9644 \text{ sau } 96{,}44\%. \quad (XV, 4.2\,b)$$

Fiecare din variantele de mai sus are avantajele și dezavantajele ei: statisticianul va folosi o variantă sau alta în funcție de scopul și sarcinile concrete ale analizei.

Indicele reducerii efective a costului producției marfă comparabile, sub influența reducerii prețului de cost, se calculează cu ajutorul formulei obișnuite (XV, 3.4):

$$I_{z(q_1)} = \frac{\Sigma z_1 q_1}{\Sigma z_0 q_1} = \frac{113\,600}{134\,000} = 0{,}8477 \text{ sau } 84{,}77\%. \quad (XV, 4.3)$$

Acest indice arată că, în medie, costul producției marfă comparabile, sub influența reducerii prețurilor efective față de cele din perioada precedentă, s-a redus cu 15,23%, realizîndu-se o economie efectivă de

$$\Sigma z_1 q_1 - \Sigma z_0 q_1 = 113\,600 - 134\,000 = -20\,400.$$

Comparînd suma economiilor obținute efectiv cu suma economiilor planificate, se observă realizarea unui volum important de economii peste plan de 17 660 lei.

Pentru scopuri analitice se poate calcula și un alt indice al dinamicii relative a costului producției marfă sub influența reducerii prețului de

cost al diferitelor produse, indice de tipul B (XV, 3,5) pentru caracteristica calitativă

$$I_{z(q_0)} = \frac{\Sigma z_1 q_0}{\Sigma z_0 q_0} = \frac{69\,200}{79\,000} = 0{,}8759 \text{ sau } 87{,}59\ \%. \qquad \text{(XV, 4.4)}$$

Este evident că un asemenea indice va fi folosit pentru a înlătura influența schimbării asortimentului producției. După cum se vede în exemplul considerat, acest indice arată o reducere mai mică a costului producției sub influența reducerii prețului de cost al diferitelor produse (12,41%) decît indicele (XV, 4.3).

Diferența de mărime a celor doi indici se explică deci numai prin modificarea asortimentului.

D. Savinski *) propune ca influența schimbării asortimentului să fie relevată prin calculul următorului coeficient :

$$k = \frac{\Sigma z_1 q_1}{\Sigma z_0 q_1} : \frac{\Sigma z_1 q_0}{\Sigma z_0 q_0}. \qquad \text{(XV, 4.5)}$$

În exemplul nostru el va fi egal cu

$$k = \frac{0{,}8477}{0{,}8759} = 0{,}9678 = 96{,}78\%.$$

Dacă la dispoziția noastră s-ar afla numai date asupra nivelului prețurilor individuale din ambele perioade și asupra producției din perioada curentă, dinamica relativă a costului producției marfă sub influența reducerii prețului de cost o putem calcula cu ajutorul formulelor transformate ale indicilor agregați, media armonică sau media aritmetică.

Se obține astfel

$$I_z(q_1) = \frac{\Sigma z_1 q_1}{\Sigma \frac{1}{i_z} z_1 q_1} = \frac{113\,600}{\left(\frac{1}{0{,}88} \cdot 88\,000\right) + \left(\frac{1}{0{,}96} \cdot 9\,600\right) + \left(\frac{1}{66} \cdot 6\,000\right)} =$$

$$= \frac{113\,600}{134\,000} = 0{,}8477 \text{ sau } 84{,}77\%$$

sau

$$I_z(q_0) = \frac{\Sigma i_z z_1 q_0}{\Sigma z_1 q_0} = \frac{(0{,}88.44\,000) + (0{,}96.19\,200) + (0{,}66.600)}{44\,000 + 19\,200 + 6\,000} =$$

$$= 0{,}8477 \text{ sau } 84{,}77\%.$$

În practică se folosește adeseori forma mediei armonice, datele necesare aflîndu-se cu ușurință în evidențele întreprinderilor.

*) Vezi în acest sens : D. Savinski, *Curs de statistică industrială*, Ed. de Stat, București, 1951, p. 330.

Pentru calculul dinamicii relative a costului producției pe ansamblul unei ramuri industriale sînt folosiți indicii de tipul C.

Să considerăm exemplul din tabelul 64.

Tabelul 64

Nr. întreprinderilor și produsele (în tone)	Perioada de bază		Perioada curentă		Costul total al producției din perioada curentă în prețul de cost	
	z_0	q_0	z_1	q_1	al perioadei de bază	al perioadei curente
					$z_0 q_1$	$z_1 q_1$
0	1	2	3	4	5	6
Întreprinderea 1						
A	25	100	20	600	15 000	12 000
B	35	60	25	400	14 000	10 000
Total parțial (*a*)	—	—	—	—	29 000	22 000
Întreprinerea 2						
A	50	400	50	500	25 000	25 000
B	85	500	65	725	61 625	47 125
Total parțial (*b*)	—	—	—	—	86 625	72 125
Total general ($a + b$)	—	—	—	—	115 625	94 125

În acest tabel sînt reprezentate datele necesare pentru calculul reducerii costului producției la un trust cu două întreprinderi; producția este comparabilă.

Indicele prețului de cost la prima întreprindere este:

$$I_{z1}(q_1) = \frac{\Sigma z_1 q_1}{\Sigma z_0 q_1} = \frac{22\,000}{29\,000} = 0{,}7586 \quad \text{sau} \quad 75{,}86\%,$$

ceea ce înseamnă că, sub influența reducerii prețului de cost al celor două produse, costul total al producției s-a redus, în medie, cu 24,14%.

Același indice la o a doua intreprindere este egal cu:

$$I_{z\,II}(q_1) = \frac{\Sigma z_1 q_1}{\Sigma z_0 q_1} = \frac{72\,125}{86\,625} = 0{,}8324 \text{ sau } 83{,}24\%,$$

deci costul total al producției s-a redus cu 16,76%, sub aceeași influență a reducerii prețului de cost al celor două produse și la această întreprindere.

Calculînd dinamica relativă medie a costului producției pe ansamblul trustului obținem

$$I_{z(I+II)}(q_1) = \frac{22\,000 + 72\,125}{29\,000 + 86\,625} = \frac{94\,125}{115\,625} = 0{,}8140 \text{ sau } 81{,}40\%. \quad (XV, 4.6)$$

După cum se vede, pe ansamblul trustului, costul producției s-a redus cu 18,60% și aceasta sub influența reducerii prețului de cost al celor două produse la ambele întreprinderi.

Pînă acum, pentru calculul dinamicii medii a costului producției pe ansamblul trustului am folosit indicii agregați de tipul B (această metodă de calcul unii autori o denumesc *metoda de întreprindere*, dat fiind că specificul ei constă în însumarea costului producției — real și convențional — obținut în fiecare întreprindere în parte).

Să aplicăm acum indicii de tipul C (adică *metoda de ramură* de calcul al dinamicii costului producției, cum se spune în literatura de specialitate) și apoi să comparăm rezultatele obținute. În acest scop trebuie calculat prețul de cost mediu al fiecărui produs în parte, pe ansamblul trustului în perioada de bază și cea curentă.

Avem:

$$\bar{z}_{0\,A(q_0)} = \frac{(25.100)+(50.400)}{500} = \frac{2\,500 + 20\,000}{500} = \frac{22\,500}{500} = 45 \text{ lei}$$

$$\bar{z}_{1\,A(q_1)} = \frac{(20.600)+(50.500)}{600 + 500} = \frac{12\,000 + 25\,000}{1\,100} = \frac{37\,000}{1\,100} = 33{,}63.$$

Aplicînd formula (XV, 3,26), obținem indicele cu structură variabilă sau indicele prețului de cost mediu al unei tone de produs A, pentru cele două perioade:

$$I\bar{z}_A = \frac{\bar{z}_1}{\bar{z}_0} = \frac{33{,}63}{45} = 0{,}7473 \text{ sau } 74{,}73\%.$$

După cum se vede, în perioada curentă prețul de cost mediu la produsul A (o tonă) s-a redus pe ansamblul trustului cu 25,27% sau în valoare absolută cu $45 - 33{,}63 = 11{,}37$ lei.

Făcînd calculul analog pentru produsul B, obținem

$$\bar{z}_{0\,B(q_0)} = \frac{44\,600}{560} = 79{,}64 \text{ lei},$$

$$\bar{z}_{1\,B(q_1)} = \frac{57\,125}{1\,125} = 50{,}77 \text{ lei},$$

iar indicele mediei pentru produsul B:

$$I\bar{z}_B = \frac{50{,}77}{79{,}64} = 0{,}6374 = 63{,}74\%.$$

Deci, prețul de cost mediu la produsul B (pe o tonă) s-a redus pe ansamblul trustului cu 36,26% sau — în valoare absolută — cu 28,87 lei.

Așadar, prețul de cost mediu — pe ansamblul trustului — s-a redus la produsul A cu 25,27% iar la produsul B cu 36,26%, pe cînd reducerea medie pe ansamblul trustului calculată după metoda de întreprindere (XV,4.6)

este de 18,60%. O asemenea contradicție are loc datorită faptului că aplicînd prima metodă de calcul al indicelui, noi am făcut abstracție de deplasarea greutății specifice a întreprinderilor în volumul total al producției pe trust, în perioada curentă.

De schimbarea structurii producției se va ține seama numai dacă se va calcula un indice cu structură variabilă al mediilor ponderate $\bar{z}_0$ și $\bar{z}_1$, luîndu-se drept pondere producția totală pe trust a fiecărui produs din perioada curentă (Q_1).

Pentru calculul acestui indice ne vom folosi de un tabel ajutător (tabelul 65).

Tabelul 65

Producția	Perioada de bază		Perioada curentă		Costul total al producției din perioada curentă în prețul de cost mediu (lei)	
	$\bar{z}_0$	Q_0	$\bar{z}_1$	Q_1	$\bar{z}_0 Q_1$	$\bar{z}_1 Q_1$
A	45	500	33,63	1 100	49 500	37 000
B	79,64	560	50,77	1 125	89 595	57 125
Total	—	—	—	—	139 095	94 125

Indicele*) căutat este

$$I_{\bar{z}(Q_1)} = \frac{\Sigma \bar{z}_1 Q_1}{\Sigma \bar{z}_0 Q_1} = \frac{94\,125}{139\,095} = 0{,}6766 \text{ sau } 67{,}66\%. \quad \text{(XV, 4.7)}$$

Prin urmare, prețul de cost al întregii producții s-a redus în medie nu cu 18,60% — după cum se deduce din formula (XV, 2.6) — ci cu 32,34%. Analiza acestei contradicții aparente o dăm la punctul 3 al acestui capitol.

Indici ai productivității muncii. Indicatorul productivității muncii ocupă un loc central în sistemul indicatorilor economiei naționale. Acest lucru se explică prin importanța hotărîtoare, politică și economică pe care o are ridicarea necontenită a nivelului productivității muncii în lupta pentru desăvîrșirea construirii socialismului în patria noastră.

Ridicarea productivității muncii este izvorul principal al reproducției socialiste lărgite și condiția esențială pentru ridicarea nivelului de trai al oamenilor muncii.

Ritmul rapid de creștere a productivității muncii — caracteristică esențială a orînduirii socialiste — se bazează, în primul rînd, pe introducerea progresului tehnic, pe ridicarea calificării muncitorilor, îmbunătățirea folosirii utilajelor, adîncirea specializării și cooperării între întreprinderi.

„Socialismul — arată tovarășul Gh. Gheorghiu-Dej în raportul C.C. al P.M.R. la cel de-al III-lea Congres al P.M.R. — asigură un cîmp

*) Tipul unui asemenea indice cu structură variabilă îl vom mai întîlni la calculul indicilor teritoriali.

nelimitat de aplicare a cuceririlor științei și tehnicii. Pentru realizarea unui ritm înalt de creștere a producției, lărgirea continuă a activității economice și social-culturale, reducerea treptată a duratei zilei de muncă, sînt necesare un număr tot mai mare de forțe de muncă și sporirea neîncetată a productivității muncii pe baza introducerii tehnicii noi în toate ramurile economici și aplicării în practică a cuceririlor științei moderne"*).

În esență, deci, socialismul generează necesitatea obiectivă și creează posibilitatea ridicării necontenite a productivității muncii, una dintre cele mai importante condiții ale victoriei socialismului asupra capitalismului, ale construirii socialismului, așa cum a arătat marele Lenin.

Creșterea productivității muncii își găsește expresia în micșorarea cantității de muncă cheltuite pentru producția unei unități de produs sau ceea ce este același lucru, în sporirea producției fără creșterea corespunzătoare a cantității de muncă cheltuită. În acest sens acționează, în socialism, *legea economică a creșterii necontenite a productivității muncii sociale* care impune ridicarea sistematică a productivității muncii în toate ramurile de producție, economisirea permanentă atît a muncii vii cît și a muncii materializate în fiecare întreprindere și pe scara întregii economii naționale. Teoria marxistă a productivității muncii ne arată că adevărata economisire constă în economisirea timpului de muncă și că această economisire este identică cu dezvoltarea forței productive a muncii.

Din punctul de vedere statistic, nivelul productivității muncii poate fi măsurat, de aceea, prin două metode:

— *metoda directă* care scoate în evidență volumul producției executate într-o unitate de timp de lucru și

— *metoda indirectă* — inversa primei — care scoate în evidență cantitatea de timp de muncă cheltuită pentru producerea unei unități de produs (consumul specific de muncă).

Vom conveni ca în cele ce urmează să folosim denumirea de indicator al productivității muncii pentru *volumul producției creat într-o unitate de timp de lucru.*

Pentru caracterizarea dinamicii relative a productivității muncii se pun două probleme esențial deosebite:

— calculul indicelui productivității muncii pentru un singur produs, la una sau la mai multe întreprinderi;

— calculul aceluiași indice, în condițiile unei mari varietăți de produse (calculul indicelui agregat).

Vom folosi următoarea notație:

w, productivitatea muncii (volumul — în expresie naturală — al producției pe unitate de timp de muncă);

w_b, idem în expresie bănească;

q, cantitatea de produse în unități naturale;

T, cheltuielile totale de timp de muncă (om-ore, om-zile etc.).

t, consumul specific de muncă vie (cheltuielile de timp de muncă vie pentru crearea unei unități de produs).

*) Gh. Gheorghiu-Dej, *Raport la cel de-al III-lea Congres al P.M.R.*, Ed. politică, București, 1960, p. 66.

Indici ai productivității muncii de tip A și B

1°. *Primul caz — producția omogenă.* Cînd este vorba de produse de același fel, fabricate la o întreprindere sau la mai multe întreprinderi, volumul total al producției (q) poate fi exprimat în unități de măsură naturale, prin însumarea directă.

În acest caz, relația

$$q = w\,T, \qquad \text{(XV, 4.8)}$$

trebuie considerată *determinantă* pentru calculul dinamicii medii a indicatorului funcțional q și a elementelor lui componente w și T.

Din (XV, 4.8) se deduce că productivitatea muncii este egală cu

$$w = \frac{q}{T}; \qquad \text{(XV, 4.9)}$$

Dacă T este considerat drept indicator funcțional, atunci avem:

$$T = \frac{q}{w}. \qquad \text{(XV, 4.10)}$$

Este evident că relația (XV, 4.8) se menține și între indicii respectivi:

$$I_q = I_w \cdot I_T \qquad \text{(XV, 4.11)}$$

adică

$$\frac{q_1}{q_0} = \frac{w_1}{w_0} \cdot \frac{T_1}{T_0}, \qquad \text{(XV, 4.12)}$$

de unde

$$\frac{w_1}{w_0} = \frac{q_1}{q_0} : \frac{T_1}{T_0}, \qquad \text{(XV, 4.13)}$$

$$\frac{T_1}{T_0} = \frac{q_1}{q_0} : \frac{w_1}{w_0}. \qquad \text{(XV, 4.14)}$$

Relația (XV, 4.13) ne arată că indicele productivității muncii este dat de raportul dintre indicele volumului producției și indicele timpului de muncă.

Pentru a caracteriza dinamica productivității muncii medii — în cazul cînd mai multe întreprinderi fabrică același produs — se va folosi desigur formula

$$I_w = \frac{\Sigma' q_1}{\Sigma' q_0} : \frac{\Sigma' T_1}{\Sigma' T_0}. \qquad \text{(XV, 4.15)}$$

Scanned with OKEN Scanner

La calculul unui asemenea indice cu structură variabilă ne întîlnim, însă, cu o problemă interesantă asupra căreia trebuie să ne oprim :

Să considerăm exemplul din tabelul 66.

Tabelul 66

Întreprinderi	Perioada de bază			Perioada curentă			Indici ai productivității muncii	Indicii parțiali ai volumului producției
	Cantitatea produselor fabricate (mii tone)	Timpul de muncă cheltuit (mii om-ore)	Productivitatea muncii pe om-oră	Cantitatea produselor fabricate (mii tone)	Timpul de muncă cheltuit (mii om-ore)	Productivitatea muncii pe om-oră		
0	1	2	3	4	5	6	7	8
	q_0	T_0	$\frac{q_0}{T_0} = w_0$	q_1	T_1	$\frac{q_1}{T_1} = w_1$	$\frac{q_1}{T_1} : \frac{q_0}{T_0} = \frac{w_1}{w_0}$	$\frac{q_1}{q_0}$
1	150	15	10	360	30	12	1,200	2,4
2	240	30	8	135	15	9	1,126	0,5625
3	30	5	6	30	5	6	1,000	1
Total	420	50	8,4	525	50	10,5	1,25	1,25

Observăm că la primele două întreprinderi, productivitatea muncii a crescut respectiv 20% și 12,6% (vezi penultima coloană) rămînînd constantă la a treia întreprindere și cu toate acestea, pe ansamblul celor trei întreprinderi, productivitatea muncii a crescut cu 25% :

$$I_w = \frac{\Sigma' q_1}{\Sigma' T_1} : \frac{\Sigma' q_0}{\Sigma' T_0} = \frac{\overline{w}_1}{\overline{w}_0} = \frac{10,5}{8,4} = 1,25 \text{ sau } 125\%. \quad \text{(XV, 4.16)}$$

Un asemenea „paradox" are adeseori loc în practica de calcul al indicilor. Se întîmplă și cazuri inverse, cînd indicele general este mai mic decît cel mai mic indice parțial (pe întreprindere).

Aceasta nu înseamnă că formula de calcul nu este bună, însă „paradoxul" ivit trebuie să capete o anumită interpretare. D. Savinski dă următoarea explicație acestui paradox :

Faptul că indicele general este mai mare decît cel mai mare indice parțial înseamnă că el reflectă influența unui factor suplimentar, în afară de creșterea nivelului productivității muncii ce a avut loc la primele două întreprinderi în parte. Este vorba și de schimbarea greutății specifice a numărului de muncitori, față de totalul muncitorilor, la prima și a doua întreprindere. La prima întreprindere la care nivelul productivității muncii a crescut cu 20%, greutatea specifică a muncitorilor a crescut și ea cu 30% (de la 30% la 60%), iar la întreprinderea a doua greutatea specifică a muncitorilor a scăzut cu 30% (de la 60% la 30%).

Prin urmare, indicele general (XV, 4.10) bazat pe compararea mediilor este, după cum știm, un *indice cu structură variabilă*, cu toate caracteristicile unui asemenea indice.

Paradoxul aparent se înlătură dacă indicele general se calculează ca medie a celor trei indici parțiali, în care drept ponderi pentru ambele perioade sînt luate aceleași cheltuieli de timp de muncă.

În literatura de specialitate se recomandă de obicei ca drept pondere să fie luați indicatorii volumului cheltuielilor de timp de muncă din perioada curentă, construindu-se *indicele cu structură* fixă al productivității muncii, de forma*)

$$I_w = \frac{\sum \left(\frac{q_1}{T_1} : \frac{q_0}{T_0} \right) T_1}{\sum T_1} = \frac{\sum i_w T_1}{\sum T_1}. \qquad \text{(XV, 4.17)}$$

În exemplul nostru avem

$$I_w = \frac{1{,}200 \cdot 30 + 1{,}126 \cdot 15 + 1 \cdot 5}{30 + 15 + 5} = \frac{57{,}90}{50} \; 1{,}1580 \text{ sau } 115{,}8\%.$$

După cum se vede, indicele obținut se plasează între cel mai mic (100%) și cel mai mare (120%) indice parțial.

Formulei (XV, 4.17) i se aduc de obicei unele transformări elementare pentru a o aduce la forma indicelui agregat — formă generală — exprimată pe baza indicatorului invers al productivității muncii.

D. V. Savinski *) și alți autori constată, în acest scop, că volumul total al cheltuielilor de timp de muncă este $T = qt$ și că substituind în formula (XV, 4.17) pe T_1 și T_0 cu valorile lor transformate $q_1 t_1$ și $q_0 t_0$ se obține următorul indice:

$$I_w = \frac{\sum \left(\frac{q_1}{q_1 t_1} : \frac{q_0}{q_0 t_0} \right) q_1 t_1}{\sum q_1 t_1} = \frac{\sum \frac{q_1 q_0 t_0}{q_1 t_1 q_0} q_1 t_1}{\sum q_1 t_1} = \frac{\sum q_1 t_0}{\sum q_1 t_1}. \qquad \text{(XV, 4.18)}$$

De sensul economic al acestei formule ne vom ocupa mai jos.

Ne vom opri acum pentru moment atrăgînd atenția asupra unei probleme metodologice principiale. Credem că autorii care recomandă indicele cu structură fixă de forma mediei aritmetice (XV, 4.17) sau cel de forma transformată (XV, 4.18) trec, în mod nejustificat, din domeniul indicilor a cărui funcție determinantă este dată de relația (XV, 4.12) și în care deci indicatorul funcțional, după cum am văzut, este *producția* omogenă, indiferent dacă se produce la o întreprindere sau la mai multe întreprinderi (ca în exemplul considerat — vezi tabelul 66), în domeniul indicelui productivității muncii ca element constitutiv al unei alte funcții determinante și anume, *volumul cheltuielilor de timp de muncă*.

*) Vezi în acest sens: D. V. Savinski, *Kurs promîșlenoi statistiki*. Gosstatizdat, Moskva, 1954, p. 171—172; T. I. Kozlov, V. Ovsienko, D. Savinski, V. Smirnski, *Kurs obșcei teorii statistiki*, 1956 p. 217—218; A. I. Gozulov, *Ekonomiceskaia statistiki*. Gostinizdat, Moskva, 1953, și alți autori. Vom vedea însă că în această formă indicele cu structură fixă nu este recomandabil (V. Peregudov), pentru că în fond aici este vorba de *structura producției* și nu a timpului de lucru. Vezi în acest sens explicația dată pe marginea formulei (XV, 3.4 a) de mai sus.

Să explicăm:

În cazul cînd un anumit produs, omogen prin natura lui (cărbune, ţiţei, oţel etc.) se produce la mai multe întreprinderi cu productivităţi diferite şi în volum diferit, se pune în mod firesc problema de a fi calculată dinamica relativă a volumului producţiei atît sub influenţa creşterii productivităţii muncii cît şi sub influenţa variaţiei volumului de timp de muncă cheltuit, faţă de perioada de bază. Am accentuat că este vorba de *dinamica relativă a producţiei*, întrucît în relaţia

$$q = w \cdot T$$

dinamica fenomenului complex este direct proporţională cu dinamica celor două elemente componente, w şi T.

Să vedem la ce fel de sistem de indici ajungem pornind de la relaţia determinantă (XV, 4.12)

$$\frac{\Sigma' q_1}{\Sigma' q_0} = \frac{\Sigma w_1 T_1}{\Sigma w_0 T_0}. \qquad \text{(XV, 4.19)}$$

Acest indice de tip A (XV, 3.3) îl putem descompune în doi indici de tip B (XV, 3.4).

Vom avea pentru caracteristica calitativă — productivitatea muncii (w) — următorul indice cu structură fixă:

$$I_w = \frac{\Sigma w_1 T_1}{\Sigma w_0 T_1}, \qquad \text{(XV, 4.20)}$$

iar pentru caracteristica de volum — cantitativă (T) — un alt indice în care drept pondere figurează caracteristica calitativă (w), la nivelul ei din perioada de bază:

$$I_T = \frac{\Sigma w_0 T_1}{\Sigma w_0 T_0}. \qquad \text{(XV, 4.21)}$$

Se vede că sistemul se verifică, întrucît

$$\frac{\Sigma w_1 T_1}{\Sigma w_0 T_0} = \frac{\Sigma w_1 T_1}{\Sigma w_0 T_1} \cdot \frac{\Sigma w_0 T_1}{\Sigma w_0 T_0}.$$

În practică putem folosi şi formulele transformate ale dinamicii, de tipul mediei armonice sau aritmetice (XV, 3.8) şi (XV, 3.9).

Astfel putem folosi formulele

$$I_w = \frac{\Sigma w_1 T_1}{\Sigma w_0 T_1} = \frac{\Sigma w_1 T_1}{\sum \frac{1}{i_w} w_1 T_1} \qquad \text{(XV, 4.22)}$$

iar pentru indicele corespunzător al timpului cheltuit,

$$I_T = \frac{\Sigma w_0 T_1}{\Sigma w_0 T_0} = \frac{\Sigma i_T w_0 T_0}{\Sigma w_0 T_0} \qquad \text{(XV, 4.23)}$$

sau și alte formule în funcție de datele pe care le avem la dispoziție; de pildă, pentru I_w se poate folosi și formula

$$I_w = \frac{\Sigma w_1 T_1}{\Sigma w_0 T_1} = \frac{\Sigma i_w w_0 T_1}{\Sigma w_0 T_1}. \qquad \text{(XV, 4.22 a)}$$

Potrivit principiului general care rezultă din funcția determinantă (XV, 4.12), ponderarea indicilor individuali trebuie făcută cu ajutorul volumului producției.

Un alt sistem de indici ce rezultă din funcția determinantă (XV, 4.12) este următorul:

$$I_w = \frac{\Sigma w_1 T_0}{\Sigma w_0 T_0},$$

$$I_T = \frac{\Sigma w_1 T_1}{\Sigma w_1 T_0},$$

$$I_{wT} = \frac{\Sigma w_1 T_1}{\Sigma w_0 T_0}.$$

După cum vom vedea și indicii din acest sistem își găsesc aplicații în practică *); de asemenea, își găsesc aplicația și formele lor transformate (forme de medii).

În exemplul nostru (tabelul 63) avem:

a) Dinamica relativă medie a volumului producției

$$I_q = \frac{\Sigma' q_1}{\Sigma' q_0} = \frac{525}{420} \cong 1{,}25 \text{ sau } 125\%. \qquad \text{(XV, 4.24)}$$

Acest indice ne arată că volumul fizic al producției a crescut sub influența schimbării concomitente pe întreprinderi a productivității muncii și a volumului de timp de muncă cheltuit, cu 25%.

Reținem atenția asupra faptului că în medie, pe ansamblul celor trei întreprinderi, indicele volumului producției este identic cu indicele productivităților medii ale muncii din cele două perioade (vezi rezultatul obținut la (XV, 4.16)).

Acest lucru se explică pe deplin dacă ținem seama că din relația (XV, 4.11) rezultă și relația

$$I_{\bar{q}} = I_{\bar{w}} \cdot I_{\bar{T}} \qquad \text{(XV, 4.25)}$$

și că în exemplul nostru concret ea devine

$$1{,}25 = 1{,}25.1,$$

*) Vezi în acest sens formula (XV, 3.5) care exprimă indicele productivității muncii cu structură fixă $\frac{\Sigma w_1 T_0}{\Sigma w_0 T_0}$ folosit la analiza indicelui cu structură variabilă de tipul $\frac{\bar{w}_1}{\bar{w}_0}$.

întrucît

$$\Sigma T_1 = \Sigma T_0 = 50 \text{ și } I_{\bar{T}} = \frac{50}{50} = 1.$$

Este evident că dacă această din urmă condiție nu este îndeplinită, vom avea $I_{\bar{q}} \neq I_{\bar{w}}$.

b) Dinamica relativă medie a volumului fizic al producției sub influența creșterii productivității muncii

$$I_w = \frac{\Sigma w_1 T_1}{\Sigma w_0 T_1} = \frac{525}{450} = 1{,}1666 \text{ sau } 116{,}66\%. \qquad \text{(XV, 4.26)}$$

Observăm că rezultatul obținut aici diferă de cel dat la (XV, 4,17). Explicația deosebirii numerice dintre cei doi indici cu structura fixă va fi dată în capitolul XVII, cu ajutorul aplicațiilor coeficientului de corelație liniară în teoria indicilor.

c) Dinamica relativă medie a volumului producției sub influența *redistribuirii* în cadrul întreprinderilor — de la o perioadă la alta — a volumului de cheltuieli de muncă:

$$I_T = \frac{\Sigma w_0 T_1}{\Sigma w_0 T_0} = \frac{450}{420} \cong 1{,}07 \text{ sau } 107\%. \qquad \text{(XV, 4.27)}$$

Sistemul de indici se verifică:

$$1{,}25 \cong 1{,}1666 \cdot 1{,}07$$

Este ușor de controlat că indicele dat de formula (XV, 4.20) este identic cu cel ce se poate obține prin formulele (XV, 4.22) și (XV, 4.22 a). De asemenea indicele dat de formula (XV, 4.21) este identic cu cel din formula (XV, 4.23).

Să revenim acum la indicele productivității muncii cu structură variabilă (XV, 4.15) și la indicele productivității cu structură fixă (XV, 4.17), implicit la forma lui transformată (XV, 4.18), relevînd caracterul acestor indici și conținutul lor economic. După părerea noastră, acești indici își găsesc locul numai în cadrul sistemului de indici în care indicatorul funcțional, determinant, este *volumul cheltuielilor de timp de muncă*. Într-adevăr, acest indicator nu poate fi ignorat în momentul în care se pune problema măsurării productivității muncii în procesul producției unor *produse eterogene* (care nu se pot însuma direct) pentru simplul motiv că în acest caz o întreprindere oarecare nu mai poate fi considerată drept unitate a colectivității de întreprinderi studiate, întrucît una și aceeași întreprindere poate să fabrice produse diferite; în această calitate va apare colectivul de muncitori (sau întreprinderea specializată) care fabrică un anumit gen de produse.

Prin urmare, colectivitatea studiată o vor constitui aceste colective de muncitori (întreprinderi specializate), fiecare unitate a colectivității avînd caracteristici: felul produsului, productivitatea muncii (w) numărul de om-ore lucrate (T), volumul total al producției (q).

Scanned with OKEN Scanner

Este evident că nici productivitatea muncii, nici volumul producției nici felul produsului nu pot fi considerați drept indicatori funcționali; rămîne, deci, numai volumul cheltuielilor de timp de muncă (numărul de om-ore) care îndeplinește condițiile unui asemenea indicator*).

Anticipînd aici, vom preciza sensul și conținutul indicilor cu structură variabilă (XV, 4.15) și cu structură fixă (XV, 4.17), relevat în majoritatea lucrărilor de specialitate.

Indicele cu structură variabilă

$$I_w = \frac{\Sigma' q'_1}{\Sigma' q_0} \cdot \frac{\Sigma' T_1}{\Sigma T_0}$$

are un sens bine determinat, exprimînd influența a doi factori în dinamică:

a) schimbări, de la o perioadă la alta, care au loc în *nivelul productivității muncii*, în diferite sectoare de muncă;

b) schimbări — în același sens — în ceea ce privește *raportul* dintre sectoarele luate în studiu; acest factor oglindește schimbarea productivității sociale medii a muncii, întrucît creșterea acestei productivități are loc atît datorită creșterii nivelului ei în fiecare sector de muncă, cît și datorită creșterii greutății specifice a întreprinderilor cu un nivel înalt al productivității muncii.

Prin urmare, indicele cu structură variabilă, prin funcția sa analitică, ne dă răspuns la întrebarea în ce măsură s-a schimbat nivelul mediu al productivității muncii pe ansamblul economiei naționale pe diferite sectoare ale muncii productive.

Acest considerent pledează pentru a acorda indicelui cu structură variabilă atributul de indice fundamental în munca de analiză statistică a creșterii productivității muncii pe ansamblul economiei naționale.

Cu ajutorul acestui indice se poate determina acea economie de timp care se realizează prin ridicarea productivității muncii în diferite sectoare ale producției materiale ca și prin redistribuirea forței de muncă în cadrul diferitelor întreprinderi sau ramuri ale economiei naționale și îndeosebi, în cadrul diferitelor ramuri ale industriei.

Trebuie să subliniem că, în cadrul măreților noastre planuri de dezvoltare în perspectivă a economiei naționale, sarcinile privind ridicarea necontenită, în ritm sporit, a productivității muncii sînt concepute și formulate, sub forma indicilor cu structură variabilă.

Indicele cu structură fixă are și el o funcție bine determinată și anume, de a elimina influența schimbării greutății specifice a diferitelor sectoare de muncă. Importanța acestui indice este determinată de faptul că schimbarea nivelului productivității muncii în diferite sectoare de muncă constituie factorul principal, primordial, în continua creștere a productivității sociale medii a muncii.

Devine limpede, deci, că în procesul muncii de analiză statistică la nivel superior, de pildă, pe ansamblul economiei naționale, indicele cu structură variabilă va servi drept instrument principal pentru relevarea

*) Vezi în acest sens: V. N. Nekraș, *Op. cit.*, p. 298.

legităților care au loc în dinamica productivității muncii, în schimb, indicele cu structură fixă va avea o funcție analitică deosebit de importantă la nivelul întreprinderilor și chiar mai mult, la nivelul ramurilor economiei naționale. Rolul acestuia din urmă rămîne însă, ajutător, relevîndu-și calitățile, îndeosebi, în analizele de detaliu pentru a determina și caracteriza statistic factorul care are rolul hotărîtor în dialectica procesului legic al creșterii productivității muncii, și anume *nivelul* ei în diferite sectoare de producție.

2°. *Al doilea caz — producția este eterogenă.* În condițiile unei mari varietăți de produse, volumul total al producției nu se poate obține prin însumarea directă a diferitelor produse deoarece ele sînt exprimate în unități de măsură naturale diferite (bucăți, tone, metri etc.)

Pentru a depăși dificultățile legate de însumarea producției, s-au elaborat patru metode principale pentru calculul dinamicii productivității muncii: metoda unităților natural-convenționale, metoda unităților de timp de muncă efectivă, metoda unităților de timp normate de muncă și metoda unităților de măsură bănești (i se spune, de obicei, metoda valorică).

Vom arăta pe scurt particularitățile fiecărei metode în parte.

a) *Metoda unităților natural-convenționale.* În esență, această metodă nu se deosebește de metoda unităților naturale decît prin aceea că în formula de calcul, indicatorul producției este exprimat în unități de măsură natural-convenționale. Cînd am examinat problema indicilor volumului producției am arătat că produsele exprimate în unități de măsură naturale diferite pot fi însumate prin folosirea unor unități de măsură-natural-convenționale (de exemplu, tractoare convenționale, hantri etc.), adică în unități ale unui produs luat convențional drept unitate de măsură. Acest lucru se poate realiza, îndeosebi, cînd este vorba de produse ce se apropie din punctul de vedere al valorii lor de întrebuințare (sau al altei caracteristici) și care se fabrică în cadrul unei anumite ramuri de producție.

Adeseori metoda în unități naturale și natural-convenționale se folosește simultan pentru a releva mărimea indicelui productivității muncii sub două aspecte și anume, cînd se ține seama de deosebirile dintre diferite tipuri de produse și cînd nu se ține seama de ele.

G. Baklanov dă exemplul din tabelul 67 *) referindu-se la producția de fontă pe un muncitor, calculată în unități naturale și paralel, în unități convenționale de fontă afînată (datele sînt pe ansamblul U.R.S.S.).

Tabelul 67

Indicatori	Anii		Indici (%) 1936=100
	1936	1937	
Producția medie de fontă pe un muncitor (în tone)	640	756	118,2
Aceeași producție exprimată în unități convenționale de fontă afînată	676	801	118,6

*) G. Baklanov, *Promîșlenaia Statistika*. Gosstatizdat, Moskva, 1953, p. 198.

Scanned with OKEN Scanner

Datele din tabel ne ilustrează atît creșterea generală a productivității muncii cît și creșterea greutății specifice a fontei de calitate mai bună, dat fiind că productivitatea medie în unități naturale (fără considerarea diferitelor calități de fontă) arată o creștere mai mică decît productivitatea medie recalculată în unități convenționale.

Pentru calculul indicelui productivității muncii în unități natural-convenționale se folosește formula recomandată în literatura de specialitate (XV, 4.17):

$$I_w = \frac{\sum \left(\frac{q_1 k}{T_1} : \frac{q_0 k}{T_0} \right) T_1}{\Sigma T_1} = \frac{\Sigma Q_1 k}{\Sigma T_1} : \frac{\Sigma Q_0 k}{\Sigma T_0} ,$$

în care qk este exprimat în unități de măsură natural-convenționale (k fiind coeficientul de transformare).

b) *Metoda unităților de timp de muncă efectivă.* Potrivit acestei metode, dinamica medie se calculează pe baza valorilor inverse ale productivității muncii, adică pe baza consumului specific de timp de muncă t, pe unitate de produs.

Am anticipat mai sus considerînd că în cazul unei producții eterogene funcția determinantă pentru calculul dinamicii relative a productivității muncii trebuie să capete, în mod necesar, următoarea formă:

$$T = \frac{q}{w} , \qquad \text{(XV, 4.28)}$$

în care T este indicatorul funcțional, reprezentînd volumul total al cheltuielilor de timp de muncă (exprimat, de obicei, în om-ore, om zile, om-ani etc.).

Privită sub aspect dinamic, relația determinantă (XV, 4.28) poate fi scrisă:

$$I_T = \frac{I_q}{I_w} = \frac{\Sigma T_1}{\Sigma T_0} = \frac{\sum \frac{q_1}{w_1}}{\sum \frac{q_0}{w_0}} , \qquad \text{(XV, 4.29)}$$

care exprimă dinamica relativă a indicatorului funcțional. Relația (XV, 4.29) ne permite să calculăm și indicii elementelor componente ale indicatorului funcțional și anume *indicele productivității muncii* (I_w) *și indicele volumului producției* (I_q).

Pentru deducerea formulelor respective cu ajutorul relației (XV, 4.29) vom raționa în felul următor *):

Din relația (XV, 4.29) rezultă că indicele productivității muncii va reprezenta dinamica medie *convențională* a volumului de cheltuieli de timp

*) Vezi A. Nekraș, *Op. cit.*

de muncă, presupunîndu-se *constant* volumul producției. Dar, presupunînd că volumul producției rămîne constant de la o perioadă la alta, *trebuie să admitem că, în acest caz, creșterea productivității muncii înseamnă reducerea cheltuielilor de timp de muncă* și invers, *creșterea cheltuielilor de timp de muncă înseamnă scăderea productivității muncii.*

Prin urmare, admițînd ipoteza că volumul producției este același în ambele perioade comparate, pentru a răspunde la întrebarea de cîte ori a crescut productivitatea muncii trebuie să stabilim, în fond, de cîte ori s-a redus volumul cheltuielilor de timp de muncă.

Aceasta înseamnă, evident, că pentru calculul indicelui productivității muncii ne vom folosi de valoarea inversă a indicelui cheltuielilor de timp de muncă, avînd grijă de a lua în calcul volumul producției din *perioada curentă*, dat fiind că se pune problema de a fi construit indicele unei caracteristici calitative.

Vom avea deci

$$I_w = \frac{\sum \frac{q_1}{w_0}}{\sum \frac{q_1}{w_1}} . \qquad \text{(XV, 4.30)}$$

Cunoscînd că productivitatea muncii (w) este invers proporțională cu volumul cheltuielilor de timp de muncă pe unitate de produs (t) și că deci avem relațiile

$$w = \frac{1}{t}, \qquad t = \frac{1}{w},$$

formula (XV, 4.30) poate fi scrisă sub o altă formă și anume:

$$I_w = \frac{\sum \frac{q_1}{\frac{1}{t_0}}}{\sum \frac{q_1}{\frac{1}{t_1}}} = \frac{\Sigma t_0 q_1}{\Sigma t_1 q_1} . \qquad \text{(XV, 4.30a)}$$

În această formă transformată, am obținut cunoscutul indice al productivității muncii, exprimat în unități de timp de muncă efectiv cheltuită.

După cum se vede, el este identic cu cel exprimat prin formula (XV, 4.18), însă deducerea lui în condițiile problemei puse aici, o considerăm ca fiind singura corectă.

Scanned with OKEN Scanner

În limitele relației noastre determinante (XV, 4.28), indicele volumului total al producției se calculează ca o dinamică convențională medie a cheltuielilor de timp de muncă, presupunînd constant consumul specific de timp de muncă (la nivelul lui din perioada de bază). Ținînd seama de aceasta ca și de faptul că la construirea indicelui productivității muncii am considerat în calcul volumul cheltuielilor de timp necesare producției din perioada curentă, indicele volumului producției (I_q) va fi exprimat prin formula

$$I_q = \frac{\sum \frac{q_1}{w_0}}{\sum \frac{q_0}{w_0}} \qquad \text{(XV, 4.31)}$$

sau, în forma ei transformată,[1]

$$I_q = \frac{\sum \frac{q_1}{\frac{1}{t_0}}}{\sum \frac{q_0}{\frac{1}{t_0}}} = \frac{\Sigma t_0 q_1}{\Sigma t_0 q_0}, \qquad \text{(XV, 4.31a)}$$

Formulele (XV, 4.29), (XV, 4.30 a) și (XV, 4.31 a) constituie un sistem la baza căruia am văzut că stă o concepție metodologică unitară.

Sistemul se verifică în cadrul categoriilor relației determinante (XV, 4.29)

$$I_T = \frac{\Sigma T_1}{\Sigma T_0} = \frac{\sum \frac{q_1}{w_1}}{\sum \frac{q_0}{w_0}} = \frac{\Sigma t_1 q_1}{\Sigma t_0 q_0} = \frac{\frac{\Sigma t_0 q_1}{\Sigma t_0 q_0}}{\frac{\Sigma t_0 q_1}{\Sigma t_1 q_1}} = \frac{\Sigma t_1 q_1}{\Sigma t_0 q_0}.$$

Indicii (XV, 4.30 a) și (XV, 4.31 a) putem să-i calculăm și sub forma mediilor aritmetice ponderate ale indicilor individuali:

$$I_w = \frac{\sum \frac{q_1}{w_0}}{\sum \frac{q_1}{w_1}} = \frac{\sum \frac{w_1}{w_0} \cdot \frac{q_1}{w_1}}{\sum \frac{q_1}{w_1}} = \frac{\Sigma i_w T_1}{\Sigma T_1}, \qquad \text{(XV, 4.32)}$$

$$I_q = \frac{\sum \frac{q_1}{w_0}}{\sum \frac{q_0}{w_0}} = \frac{\sum \frac{q_1}{q_0} \cdot \frac{q_0}{w_0}}{\sum \frac{q_0}{w_0}} = \frac{\Sigma i_q T_0}{\Sigma T_0}. \qquad \text{(XV, 4.33)}$$

Să ilustrăm calculul indicilor din sistemul considerat cu ajutorul datelor din tabelul 68.

Tabelul 68

Produse	Perioada de bază				Perioada curentă				Indici individuali			
	Volumul producției, în buc.	Volumul cheltuielilor de timp muncă (om-zile)	Productivitatea muncii coloana 1 / coloana 2	Consum specific de timp de muncă coloana 2 / coloana 1	Volumul producției, în buc.	Volumul cheltuielilor de timp muncă (om-zile)	Productivitatea muncii coloana 5 / coloana 4	Consum specific de timp de muncă coloana 6 / coloana 5				
	1	2	3	4	5	6	7	8	9	10	11	12
	q_0	T_0	$\frac{q_0}{T_0} = w_0$	t_0	q_1	T_1	$\frac{q_1}{T_1} = w_1$	t_1	i_q	i_w	i_t	i_z
A	100	50	2	0,5	120	60	2	0,5	1,2	1	1	1,2
B	200	80	2,5	0,4	240	120	2	0,5	1,2	0,8	1,25	1,5
C	1 000	1 000	1	1	1 500	2 000	0,75	1,33	1,5	0,75	1,33	1,5
Total	—	1 130	—	—	—	2 180	—	—				

$$I_T = \frac{\Sigma T_1}{\Sigma T_0} = \frac{2180}{1130} = 1{,}9292 \text{ sau } 192{,}92\%,$$

$$I_w = \frac{\Sigma t_0 q_1}{\Sigma t_1 q_1} = \frac{(0{,}5 \cdot 120) + (0{,}4 \cdot 240) + (1 \cdot 500)}{(0{,}5 \cdot 120) + (0{,}5 \cdot 240) + (1{,}33 \cdot 1500)} =$$

$$= \frac{1656}{2180} = 0{,}7596 \text{ sau } 75{,}96\%$$

$$I_q = \frac{\Sigma t_0 q_1}{\Sigma t_0 q_0} = \frac{1656}{1130} = 1{,}4654 \text{ sau } 146{,}54\%. \qquad \text{(XV, 4.32a)}$$

După cum se vede, în exemplul considerat, volumul cheltuielilor de timp de muncă a crescut în perioada curentă cu 92,9%. Dacă productivitatea muncii ar fi rămas constantă, aceasta ar fi însemnat că și volumul producției ar fi crescut aproximativ cu același procent; se constată însă că volumul producției a crescut numai cu 46,54%, ceea ce înseamnă că productivitatea muncii a scăzut considerabil. Într-adevăr, indicele respectiv (I_w) arată o scădere a productivității muncii cu 24,04%.

Scanned with OKEN Scanner

Aplicînd formulele respective ale mediilor aritmetice (XV, 4.32) și (XV,4.33) ajungem la același rezultat :

$$I_w = \frac{\Sigma i_w T_1}{\Sigma T_1} = \frac{(1 \cdot 60) + (0{,}8 \cdot 120) + (0{,}75 \cdot 2000)}{2180} =$$

$$= \frac{1656}{2180} = 0{,}7596 \quad \text{sau} \quad 75{,}96\%,$$

$$I_q = \frac{\Sigma i_q T_0}{\Sigma T_0} = \frac{(1{,}2 \cdot 50) + (1{,}2 \cdot 80) + (1{,}5 \cdot 1000)}{1130} =$$

$$= \frac{1656}{1130} = 1{,}4654 \quad \text{sau} \quad 146{,}54\%.$$

Dacă, folosind aceleași date, ne întrebăm care este dinamica medie a consumului specific de timp de muncă, răspunsul îl putem da în două variante — în ambele cu același rezultat.

Prima variantă. Raționăm în felul următor :

Dat fiind că avem relațiile $w = \frac{1}{t}$ și $t = \frac{1}{w}$, ele se vor păstra și în dinamică.

Vom avea prin urmare

$$I_w = \frac{1}{I_t} \quad \text{și} \quad I_t = \frac{1}{I_w}.$$

Egalitatea a doua ne dă răspunsul la întrebarea pusă. O putem scrie sub forma următoare :

$$I_t = \frac{1}{I_w} = \frac{1}{\frac{\Sigma t_0 q_1}{\Sigma t_1 q_1}} = \frac{\Sigma t_1 q_1}{\Sigma t_0 q_1}. \qquad \text{(XV, 4.34)}$$

Ea ne arată că indicele consumului specific de timp de muncă este o mărime inversă a indicelui care caracterizează dinamica medie a productivității muncii.

A doua variantă. Recurgem din nou la indicatorul funcțional T pe care îl exprimăm în funcție de t și q.

Vom avea, în acest caz, relația determinantă de forma produsului și anume :

$$T = t \cdot q. \qquad \text{(XV, 4.35)}$$

Pe baza principiilor de ponderare pe care le-am formulat, dinamica mediei a consumului specific de timp de muncă (t) este caracterizată printr-un indice de tip B (XV, 3.4).

Avem deci

$$I_t(q_1) = \frac{\Sigma t_1 q_1}{\Sigma t_0 q_1} \qquad \text{(XV, 4.34 a)}$$

şi

$$I_q(t_0) = \frac{\Sigma t_0 q_1}{\Sigma t_0 q_0}. \qquad \text{(XV, 4.36)}$$

Pe baza datelor din exemplul nostru (tabelul 65) avem:

$$I_t(q_1) = \frac{2\,180}{1\,656} = 1{,}3164 \quad \text{sau} \quad 131{,}64\%,$$
$$I_q(t_0) = \frac{1\,656}{1\,130} = 1{,}4654 \quad \text{sau} \quad 146{,}54\%. \qquad \text{(XV, 4.37)}$$

Relaţia determinantă se verifică:

$$I_T = \frac{\Sigma T_1}{\Sigma T_0} = \frac{\Sigma t_1 q_1}{\Sigma t_0 q_0} = \frac{\Sigma t_1 q_1}{\Sigma t_0 q_1} \cdot \frac{\Sigma t_0 q_1}{\Sigma t_0 q_0}$$

sau, concret:

$$\frac{2\,180}{1\,130} = 1{,}9292 \cong 1{,}3164 \cdot 1{,}4654.$$

Putem constata acum că o dată cu scăderea productivităţii muncii cu 24,04% (vezi rezultatul de la (XV, 4.32a)) s-a produs şi fenomenul invers, consumul specific de timp de muncă a crescut cu 31,64%.

Mai trebuie să atragem atenţia că indicele consumului specific de timp de muncă se mai poate calcula folosindu-ne de media aritmetică sau media armonică obţinîndu-se acelaşi rezultat; idem indicele volumului producţiei.

Astfel, avem:

— indicele consumului specific

$$I_t = \frac{\Sigma i_t t_0 q_1}{\Sigma t_0 q_1} = \frac{(1.0{,}5 \cdot 120) + (1{,}25 \cdot 0{,}4 \cdot 240) + (1{,}33 \cdot 1.1500)}{1\,656} \cong$$
$$\cong \frac{2\,180}{1\,656} = 1{,}3164 \quad \text{sau} \quad 131{,}64\%$$

sau

$$I_t = \frac{\Sigma t_1 q_1}{\sum \frac{1}{i_t} \cdot t_1 q_1} = \frac{2\,180}{1\,656} = 1{,}3164 \quad \text{sau} \quad 131{,}64\%;$$

— indicele volumului fizic

$$I_q = \frac{\Sigma i_q t_0 q_0}{\Sigma t_0 q_0} = \frac{\sum \frac{q_1}{q_0} t_0 q_0}{\Sigma t_0 q_0} = \frac{\Sigma t_0 q_1}{\Sigma t_0 q_0} = 1{,}4654 \quad \text{sau} \quad 146{,}54\%.$$

Scanned with OKEN Scanner

Indicele productivității muncii construit după metoda unităților de timp de muncă efectivă (XV, 4.30 a) sau (XV, 4.32) prezintă anumite particularități pe care trebuie să le scoatem în evidență.

În primul rînd, în afară de faptul că ilustrează dinamica relativă a productivității muncii, acest indice mai relevă cantitativ și ce anume volum de muncă, estimat în unitățile consumului specific de timp de muncă din perioada de bază, revine la o unitate de timp de muncă din perioada curentă. De exemplu, dacă ne referim la rezultatul obținut prin formula (XV, 4.32 a) el ne mai arată că o cheltuială de timp de muncă din perioada curentă de un om-zile revine numai 0,7596 de om-zile din perioada de bază.

În al doilea rînd, acest indice se bucură de proprietatea de a putea fi calculat numai pe baza datelor asupra volumului producției din perioada curentă (în expresie naturală) și asupra cheltuielilor de timp efective.

O altă particularitate a acestui indice constă în aceea că se calculează numai în condițiile unui *asortiment comparabil* și că ia în considerare numai timpul de muncă al muncitorilor *direct productivi*.

Pentru ca acest indice să fie aplicabil în studiul dinamicii productivității muncii tuturor muncitorilor, i se poate aduce o corecție corespunzătoare înmulțindu-l cu indicele greutății specifice a cheltuielilor de timp de muncă a muncitorilor productivi, în volumul total al cheltuielilor de timp de muncă.

Această corecție a fost propusă de D. V. Savinski și reluată de G. Baklanov *).

Notînd cu a cheltuielile de timp de muncă ale muncitorilor direct productivi și cu b cele ale muncitorilor auxiliari, vom avea

$$\frac{a_0}{a_0 + b_0} = k_0$$

greutatea specifică a cheltuielilor de timp de muncă ale muncitorilor direct productivi, în perioada de bază și

$$\frac{a_1}{a_1 + b_1} = k_1$$

același indicator, în perioada curentă.

Raportul $\frac{k_1}{k_0}$ ne dă indicele respectiv care se înmulțește cu indicele productivității muncii $\frac{\Sigma t_0 q_1}{\Sigma t_1 q_1}$, obținîndu-se dinamica productivității muncii privind totalitatea muncitorilor.

c) *Metoda unităților de timp normate de muncă*. Indicele productivității muncii calculat după această metodă introduce un element nou și anume *consumul specific de timp de muncă normat*, care, evident, este constant de la o perioadă la alta.

*) G. Baklanov, *Op. cit.*, p. 214–215.

Scanned with OKEN Scanner

Forma acestui indice va fi deci:

$$I_{t_n} = \frac{\Sigma q_1 t_n}{\Sigma q_1 t_1} : \frac{\Sigma q_0 t_n}{\Sigma q_0 t_0} \qquad \text{(XV, 4.38)}$$

în care t_n este consumul specific *normat* de timp de muncă.

Funcţia determinantă din care se poate deduce acest indice trebuie să fie volumul producţiei exprimat în unităţi de timp normat, $\Sigma q t_n$. Mai exact, $\Sigma q t_n$ exprimă *volumul activităţii de producţie.*

Indicele productivităţii muncii exprimat în unităţi de timp normat se poate obţine indirect prin împărţirea indicelui volumului activităţii de producţie la indicele cheltuielilor de timp efective:

$$\frac{\Sigma q_1 t_n}{\Sigma q_0 t_n} : \frac{\Sigma T_1}{\Sigma T_0} = \frac{\Sigma q_1 t_n}{\Sigma T_1} : \frac{\Sigma q_0 t_n}{\Sigma T_0} = \frac{\Sigma q_1 t_n}{\Sigma q_1 t_1} : \frac{\Sigma q_0 t_n}{\Sigma q_0 t_0}.$$

d) *Indici ai productivităţii muncii în expresie băneasca* *) (valorică). Pentru a înlătura dificultăţile legate de elaborarea indicilor sintetici în cazul unor colectivităţi de produse eterogene, neînsumabile direct, teoria şi practica metodei indicilor a introdus în calcul categoria *valorii* în expresia ei bănească, omogenizînd din acest punct de vedere colectivităţile studiate şi făcîndu-le comparabile.

Indice al productivităţii muncii medii (indice cu structură variabilă) de tipul C. Indicele productivităţii muncii în expresie bănească poate scoate în evidenţă, în primul rînd, dinamica productivităţii muncii medii: $\overline{w}_{b1}$ faţă de $\overline{w}_{b0}$

$$I_{w_b}{}^{**}) = \frac{\overline{w}_{b1}}{\overline{w}_{b0}} = \frac{\Sigma p q_1}{\Sigma T_1} : \frac{\Sigma p q_0}{\Sigma T_0} = \frac{\Sigma p q_1}{\Sigma p q_0} : \frac{\Sigma T_1}{\Sigma T_0}. \qquad \text{(XV, 4.39)}$$

Este vorba, evident, de indicele cu structură variabilă a dinamicii productivităţii muncii în expresie bănească.

Asupra indicilor sintetici în expresie valorică, cu structură variabilă, exercită influenţa, după cum ştim, doi factori:

a) schimbarea nivelului productivităţii muncii în diferite întreprinderi şi

b) schimbarea greutăţii specifice a întreprinderilor cu nivelul absolut al productivităţii muncii diferit.

---

*) Credem că folosirea acestei noţiuni este mai potrivită dat fiind că „banii" sînt numai una din formele valorii, iar la elaborarea aşa-numiţilor indici „valorici" (noţiune folosită de obicei în lucrările de specialitate) noi recurgem tocmai la această formă a ei — expresie bănească.

**) Trebuie să precizăm că pentru calculul productivităţii muncii în expresie valorică, indicatorii producţiei (pq) pot fi exprimaţi sub trei forme principale diferite:

a) Se consideră *producţia globală* în preţuri constante în forma ei obişnuită de calcul, stabilită prin metodologia oficială elaborată de Direcţia Centrală de Statistică de pe lîngă Consiliul de Miniştri al R.P.R.

b) Se consideră *producţia netă* exprimînd cheltuielile de timp de muncă vie din perioada curentă.

c) Se consideră *producţia totală*, producţia globală, calculată după metoda economiei naţionale care include în afară de munca vie şi amortizarea utilajului şi a altor mijloace ale muncii din perioada precedentă.

Înlăturarea influenței acestui din urmă factor asupra indicelui mediu general al productivității muncii în expresie bănească se face de obicei folosindu-se metoda recomandată de acad. S. G. Strumilin *) pentru calculul dinamicii medii.

În esență, această metodă constă în calculul unui indice mediu, în construcția căruia intră indicii medii pe întreprinderi, ponderați cu indicatorii de structură din perioada curentă.

Formula de calcul este următoarea:

$$I_{\text{pr. m.}} = \frac{\Sigma i_w T_1}{\Sigma T_1}, \qquad \text{(XV, 4.40)}$$

în care i_w este indicele mediu al productivității muncii la diferite întreprinderi. Se observă că el este de tipul indicelui dat de formula (XV, 2.17).

Să considerăm exemplul din tabelul 69.

Tabelul 69

Întreprinderi	Perioada de bază			Perioada curentă			Indicii productivității muncii
	Producția globală în prețuri constante (mii lei)	Nr. mediu scriptic al muncitorilor	Productivitatea muncii	Producția globală în prețuri constante (mii lei)	Nr. mediu scriptic al muncitorilor	Productivitatea muncii	
	$Pq_0 = W_0$	T_0	w_0	$Pq_1 = W_1$	T_1	w_1	i_w
0	1	2	3	4	5	6	7
1	37 200(31 %)	1 140(38 %)	32,63	66 000(33 %)	1 596(42 %)	41,35	1,267
2	32 400(27 %)	840(28 %)	38,57	50 000(25 %)	950(25 %)	52,63	1,364
3	50 400(48 %)	1 020(34 %)	49,41	84 000(42 %)	1 254(33 %)	66,98	1,355
Total trust	120 000(100 %)	3 000(100 %)	40,00	200 000(100 %)	3 800(100 %)	52,63	1,3157

Indicele productivității muncii calculat pe ansamblul celor trei întreprinderi (pe trust) după metoda propusă de S.G. Strumilin va fi

$$I_{\text{pr. m. str. fixă}} = \frac{\Sigma i_w T_1}{T_1} = \frac{(1{,}267 \cdot 1\,596) + (1{,}364 \cdot 950) + (1{,}355 \cdot 1\,254)}{1\,596 + 950 + 1\,254} =$$

$$= \frac{2022{,}13 + 1\,295{,}80 + 1\,699{,}17}{3\,800} = \frac{5\,017{,}10}{3\,800} = 1{,}3202 \text{ sau } 132{,}02\,\%.$$

După cum se vede, indicele general al productivității muncii (1,3157), calculat direct pe baza nivelurilor medii ale productivității muncii

*) S. G. Strumilin (n. 1877), cunoscut economist și statistician sovietic, are o bogată activitate științifică, teoretică și practică în domeniul economiei și statisticii. A scris și publicat un număr de peste 180 de cărți și articole, consacrate statisticii, evidenței, planificării socialiste și istoriei economiei naționale a U.R.S.S.

Scanned with OKEN Scanner

$\left(\frac{\Sigma q_1}{\Sigma T_1} : \frac{\Sigma q_0}{\Sigma T_0} = \frac{\overline{w}_1}{\overline{w}_0} = 1{,}3157\right)$, este ceva mai mic decît indicele calculat după metoda lui S. Strumilin. Explicația acestui fapt ce se dă de obicei de unii autori, constă în aceea că greutatea specifică a întreprinderii 1 (tabelul 69), în care nivelul productivității muncii este cel mai scăzut, a crescut în perioada curentă de la 38% la 42% (în ceea ce privește numărul scriptic al muncitorilor).

Dacă, dimpotrivă, ar fi crescut greutatea specifică a întreprinderii 3, care are cel mai mare nivel al productivității muncii, atunci, evident, s-ar fi produs fenomenul invers pe ansamblul trustului și anume: $\frac{\overline{w}_1}{\overline{w}_0}$ ar fi fost mai mare decît indicele calculat după metoda lui S. Strumilin.

Prin urmare, indicele productivității muncii calculat după metoda lui S. Strumilin elimină influența schimbării greutății specifice a întreprinderilor cu un nivel diferit al productivității muncii.

Indicele lui S. Strumilin este deci un indice cu structură fixă*), spre deosebire de indicele dat de formula (XV, 4.39), care am văzut că este un indice cu structură variabilă.

Deosebirea principială dintre acești doi indici am arătat-o atunci cînd am examinat calculul productivității muncii în ipoteza unui produs care se fabrică la mai multe întreprinderi; în speță, problema este aceeași. Indicele cu structură variabilă are un rol important în analizele statistice, îndeosebi la nivelul ramurii sau al economiei naționale.

Am constatat că indicele productivității muncii în expresie bănească calculat pe baza indicatorului direct poate avea aplicații pe scară mult mai largă decît indicele calculat după metoda unităților de timp de muncă efectivă $\left(\frac{\Sigma t_0 q_1}{\Sigma t_1 q_1}\right)$, în construcția lui intrînd elemente care se găsesc cu ușurință în cele trei forme de evidență a proceselor și fenomenelor care au loc în cadrul larg al reproducției socialiste lărgite. Din aceasta însă nu putem trage concluzia că indicele productivității muncii în expresie bănească ar avea o prioritate *teoretică* sau *principială* față de celălalt. Problema raportului dintre acești indici și a condițiilor în care valorile lor coincid este o problemă deosebită asupra căreia vom reveni.

Totuși, atragem atenția de pe acum că între cei doi indici ai productivității muncii există o anumită legătură reciprocă căreia i se poate da o anumită formă matematică precisă **).

Sistemul de indici ai dinamicii relative. Funcția determinantă care trebuie să stea la baza calculului dinamicii productivității muncii în expresie bănească este identică — prin conținutul categoriilor pe care

*) Vom vedea mai jos că pentru a analiza influența deplasărilor structurale trebuie luat în calcul un alt indice cu structură fixă și anume cel dat de formula (XV, 3.5). Menționăm aici că indicele lui S. Strumilin nu elimină influența structurii timpului de lucru ci influența *structurii producției*.

**) Vezi în acest sens L. S. Kazineț, *Uciontie zapiski po statistiki*. vol. I, Moskva, 1955. Teoria lui L. S. Kazineț este expusă în capitolul următor.

Scanned with OKEN Scanner

le cuprinde — cu cea dată de relația (XV, 4.8) care ne-a servit pentru calculul dinamicii productivității muncii în expresie naturală, în ipoteza unei producții omogene. Deosebirea constă numai în faptul că indicatorul funcțional — volumul producției — și indicatorul productivității muncii capătă expresie bănească.

Vom avea deci, funcția determinantă

$$pq = w_b T, \qquad \text{(XV, 4.41)}$$

din care deducem sistemul de indici caracterizînd :

a) dinamica relativă a producției

$$I_{pq} = \frac{\Sigma pq_1}{\Sigma pq_0}; \qquad \text{(XV, 4.42)}$$

b) dinamica relativă a producției sub influența schimbării productivității muncii

$$I_{w_b} = \frac{\Sigma w_{b_1} T_1}{\Sigma w_{b_0} T_1}; \qquad \text{(XV, 4.43)}$$

c) dinamica relativă a producției sub influența schimbării volumului de cheltuieli de timp de muncă

$$I_T = \frac{\Sigma w_{b_0} T_1}{\Sigma w_{b_0} T_0}. \qquad \text{(XV, 4.44)}$$

Sistemul se verifică

$$I_{pq} = \frac{\Sigma pq_1}{\Sigma pq_0} = \frac{\Sigma w_{b_1} T_1}{\Sigma w_{b_0} T_0} = \frac{\Sigma w_{b_1} T_1}{\Sigma w_{b_0} T_1} \cdot \frac{\Sigma w_{b_0} T_1}{\Sigma w_{b_0} T_0}.$$

În acest sistem, indicii I_{w_b} și I_T pot fi înlocuiți prin indici identici de forma mediilor aritmetice sau armonice corespunzătoare :

$$I_{w_b} = \frac{\Sigma w_{b_1} T_1}{\sum \frac{1}{i_{w_{b_1}}} w_{b_1} T_1} \quad \text{sau} \quad I w_b = \frac{\Sigma i_{w_b} w_{b_0} T_1}{\Sigma w_{b_0} T_1} \qquad \text{(XV, 4.43a)}$$

$$I_T = \frac{\Sigma i_T w_{b_0} T_0}{\Sigma w_{b_0} T_0}. \qquad \text{(XV, 4.44a)}$$

Dacă comparăm indicele volumului producției omogene (un singur fel de produs) $\frac{\Sigma' q_1}{\Sigma' q_0}$ cu indicele în expresia valorică (XV, 4.42) calculat tot pentru un singur produs, adică $\frac{\Sigma' pq_1}{\Sigma' pq_0}$, constatăm că primul nu este

decît un caz particular al celui de-al doilea. Acelaşi lucru putem spune şi în ceea ce priveşte indicele productivităţii muncii calculat în unităţi băneşti, unităţi naturale şi unităţi de timp de muncă (desigur, în ipoteza unui singur produs).

Avem

$$\frac{\Sigma' pq_1}{\Sigma' T_1} : \frac{\Sigma' pq_0}{\Sigma' T_0} = \frac{\Sigma' q_1}{\Sigma' T_1} : \frac{\Sigma' q_0}{\Sigma' T_0} = \frac{\Sigma' T_0}{\Sigma' q_0} : \frac{\Sigma' T_1}{\Sigma' q_1}$$

sau

$$\frac{\overline{w}_{b_1}}{\overline{w}_{b_0}} = \frac{\overline{w}_1}{w_0} = \frac{\bar{t}_0}{\bar{t}_1}.$$

5. Analiza indicelui cu structură variabilă

În cadrul analizei statistice se pune adeseori problema ca indicele cu structură variabilă, de tipul C (XV, 3.24), să fie descompus pe elemente componente pentru a se vedea influenţa pe care o exercită cei doi factori cunoscuţi:

— schimbarea nivelului indicatorului calitativ α de la o perioadă la alta, şi

— schimbarea structurii indicatorului de volum ψ de la o perioadă la alta.

După cum se ştie, este vorba de o problemă deosebit de importantă. De rezolvarea ei s-au ocupat mai mulţi autori, însă, după cîte se pare, ei n-au ajuns încă la soluţii unitare, problema continuînd a fi controversată.

Din punct de vedre teoretic, problema constă în a elimina din construcţia indicelui cu structură variabilă, elementul care exprimă variaţia structurii. S-a propus în acest scop ca *indicele cu structură variabilă să fie împărţit la indicele cu structură fixă*, obţinîndu-se tocmai indicele deplasărilor structurale în raport cu caracteristica calitativă studiată, de tipul aceluia dat de formula (XV, 4.26), adică

$$I_{\psi(\alpha_0)} = \frac{\Sigma \alpha_0 \psi_1^*}{\Sigma \alpha_0 \psi_0^*}.$$

Este evident că trebuie să avem, în acest caz, relaţia

$$I_{\text{str. fixă}} \cdot I_{\text{depl. str.}} = I_{\text{str. variabile}}. \qquad \text{(XV, 5.1)}$$

Pentru măsurarea statistică a deplasărilor structurale s-a adoptat în practică formula (XV, 4.26) care ţine seama de principiul de ponderare consacrat la construirea indicilor: caracteristica *calitativă* se ponderează cu indicatorii caracteristicii cantitative şi invers, caracteristica *cantitativă* se ponderează cu nivelurile caracteristicii calitative (ca în cazul indicelui (XV, 4.27)).

Scanned with OKEN Scanner

Prin urmare, relația (XV,5.1) poate fi scrisă sub forma:

$$\frac{\bar{\alpha}_1}{\bar{\alpha}_0} = \frac{\Sigma\alpha_1\psi_1}{\Sigma\alpha_0\psi_1} \cdot \frac{\Sigma\alpha_0\psi_1'}{\Sigma\alpha_0\psi_0'}. \qquad \text{(XV, 5.2)}$$

Se constată, însă, că aplicarea în toate cazurile a principiului de ponderare consacrat duce la unele contradicții din care nu se poate ieși decît schimbînd — în cazul special al analizei indicelui cu structură variabilă — sistemul de ponderare.

Vom învedera această necesitate pe exemplul concret al diferiților indici cu structură variabilă.

1°. *Analiza indicelui productivității muncii cu structură variabilă.* Să revenim din nou la datele din tabelul 69 și să prezentăm modul în care se descompune, de obicei, indicele cu structură variabilă

$$\frac{\bar{w}_1}{\bar{w}_0}.$$

Potrivit regulii cunoscute vom avea

$$I_{\text{depl. str.}} = \frac{\bar{w}_1}{\bar{w}_0} : \frac{\Sigma i_w T_1}{\Sigma T_1}$$

sau înlocuind datele concrete:

$$I_{\text{depl. str.}} = \frac{1{,}3157}{1{,}3202} = 0{,}9965 \text{ sau } 99{,}65\%.$$

Acest indice — după părerea unor autori — este menit a măsura influența pe care o exercită *redistribuirea* forței de muncă în perioada curentă, interpretarea lui fiind simplă.

Dacă avem inegalitatea

$$I_{\text{depl. str.}} > 1, \qquad \text{(XV, 5.3)}$$

înseamnă că deplasările structurale au fost *favorabile* pentru creșterea generală a productivității muncii (a crescut greutatea specifică a întreprinderilor la care nivelul productivității muncii este mai ridicat) și invers, dacă avem inegalitatea

$$I_{\text{depl. str.}} < 1, \qquad \text{(XV, 5.4)}$$

înseamnă că deplasările structurale *nu au fost favorabile* (a crescut greutatea specifică a întreprinderilor cu un nivel mai scăzut al productivității muncii).

Tocmai în cazul inegalității (XV, 5.4) ne aflăm și în exemplul considerat,

$$0{,}9965 < 1.$$

Să considerăm acum un alt exemplu asemănător (tabelul 70)

Tabelul 70

Întreprinderi	Perioada de bază			Perioada curentă			Indici
	Producția globală în prețuri constante (mii lei)	Nr. mediu scriptic al muncitorilor	Productivitatea muncii	Producția globală în prețuri constante (mii lei)	Nr. mediu scriptic al muncitorilor	Productivitatea muncii	
	$Q_0 = pq_0$	T_0	$w_0 = \frac{Q_0}{T_0}$	$Q_1 = pq_1$	T_1	$w_1 = \frac{Q_1}{T_1}$	$i_w = \frac{w_1}{w_0}$
0	1	2	3	4	5	6	7
1	37 200(31 %)	1 140(38 %)	32,63	66 000(33 %)	840(20 %)	78,57	2,40
2	32 400(27 %)	840(28 %)	38,57	50 000(25 %)	1 260(30 %)	39,68	1,02
3	50 400(42 %)	1 020(34 %)	49,41	84 000(42 %)	2 100(50 %)	40,00	0,80
Total trust	120 000(100 %)	3 000(100 %)	40,00	200 000(100 %)	4 200(100 %)	47,61	1,19

Avem

$$I_{\bar{w}} = \frac{\bar{w}_1}{\bar{w}_0} = 1{,}19 \text{ sau } 119\%$$

$$I_{\text{str. fixă}} = \frac{(2{,}40 \cdot 840) + (1{,}02 \cdot 1\,260) + (0{,}80 \cdot 2\,100)}{840 + 1\,260 + 2\,100} =$$

$$= \frac{4\,981{,}20}{4\,200} = 1{,}186 \text{ sau } 118{,}6\%,$$

$$I_{\text{depl. str.}} = \frac{1{,}19}{1{,}186} = 1{,}003. \qquad \text{(XV, 5.4a)}$$

După cum se vede, $I_{\text{depl. str.}} > 1$, ceea ce înseamnă că ar trebui să ne aflăm într-un caz „favorabil" al deplasărilor structurale pentru creșterea generală a productivității muncii, în sensul că a avut loc *redistribuirea* favorabilă a forței de muncă și că a crescut, deci, greutatea specifică a numărului de muncitori la întreprinderile unde avem în perioada curentă un nivel al productivității muncii mai ridicat. Or, în realitate a avut loc tocmai fenomenul invers celuia pe care ni-l ilustrează indicele deplasărilor structurale. Într-adevăr, se poate ușor constata că la întreprinderile 3 și 2, cu un nivel al productivității muncii scăzut, a crescut considerabil greutatea specifică a muncitorilor în numărul lor total pe ansamblul trustului, iar la întreprinderea 1 cu cel mai ridicat nivel al productivității muncii, greutatea specifică a muncitorilor a scăzut considerabil (de la 38% la 20%).

Scanned with OKEN Scanner

Avem deci o contradicție evidentă între caracterizarea ce ne-o face indicele deplasărilor structurale și realitatea ilustrată de datele exemplului.

Explicația unei asemenea contradicții a dat-o V. Peregudov *). Să urmăm raționamentul teoretic care se face de obicei în procesul determinării indicelui deplasărilor structurale, pentru a releva verigile lui îndoielnice din punct de vedere logic.

Se consideră de obicei egalitățile

$$I_{\bar{w}} = \frac{\Sigma Q_1}{\Sigma T_1} : \frac{\Sigma Q_0}{\Sigma T_0} = \frac{\Sigma Q_1 \Sigma T_0}{\Sigma Q_0 \Sigma T_1};$$

$$I_{\bar{w}}^{\text{str. fixă}} = \frac{\Sigma \left(\frac{Q_1}{T_1} : \frac{Q_0}{T_0} \right) T_1}{\Sigma T_1} = \frac{\Sigma \frac{Q_1 T_0}{Q_0}}{\Sigma T_1};$$

$$I_{\bar{w}}^{\text{schimb. str.}} = \frac{\Sigma Q_1 \Sigma T_0}{\Sigma Q_0 \Sigma T_1} : \frac{\Sigma \frac{Q_1 T_0}{Q_0}}{\Sigma T_1} = \frac{\Sigma Q_1 \Sigma T_0}{\Sigma Q_0} : \Sigma \frac{Q_1 T_0}{Q_0} = \frac{\Sigma \frac{T_0}{Q_0} Q_0}{\Sigma Q_0} : \frac{\Sigma \frac{T_0}{Q_0} Q_1}{\Sigma Q_1}.$$

Dacă ținem seama că $\frac{T_0}{Q_0}$ reprezintă t_0 — consumul specific de muncă pe unitate de produs în cadrul unei întreprinderi în perioada de bază și că

$$Q_0^* = \frac{Q_0}{\Sigma Q_0} \quad \text{și} \quad Q_1^* = \frac{Q_1}{\Sigma Q_1},$$

atunci indicele deplasărilor structurale este egal cu:

$$I_{\bar{w}}^{\text{schimb. str.}} = \frac{\Sigma \frac{T_0}{Q_0} Q_0}{\Sigma Q_0} : \frac{\Sigma \frac{T_0}{Q_0} Q_1}{\Sigma Q_1} = \frac{\Sigma t_0 Q_0}{\Sigma Q_0} : \frac{\Sigma t_0 Q_1}{\Sigma Q_1} = \frac{\Sigma t_0 Q_0^*}{\Sigma t_0 Q_1^*}.$$

În ultima sa formă, indicele deplasărilor structurale este folosit de obicei în cercetările teoretice și practice, așa după cum am mai arătat.

Am văzut, însă, că tocmai în această formă el ne-a dus la contradicția amintită. Cu privire la această formă a indicelui deplasărilor structurale se pot face următoarele observații:

a) în construcția lui nu apare volumul cheltuielilor de timp de muncă din perioada curentă (T_1);

b) deplasările structurale pe care le caracterizează nu se referă la indicatorul cheltuielilor de timp de muncă din perioada curentă (T_1), ci la indicatorul *volumului producției* (producția globală Q_1);

*) V. Peregudov, *Ucionlie zapiski po statistike*, vol. I, Moskva, 1955, p. 86—88.

c) deplasările structurale sînt estimate în raport cu consumul specific de timp de muncă din perioada de bază t_0;

d) el va fi mai mare decît 1 dacă corespondenţa dintre indicatorii de structură (Q_0^*) şi indicatorii nivelului consumului specific, de timp de muncă (t_0), se apropie mai mult de o corespondenţă proporţională, decît de corespondenţa dintre Q_1^* şi t_0; în situaţie inversă el va fi mai mic decît 1.

În ansamblu, acest indice *nu poate măsura efectul redistribuirii forţei de muncă în timp — de la perioada de bază la perioada curentă* — dat fiind că nu conţine decît datele privind cheltuielile de timp de muncă din perioada de bază.

Devine, deci, explicabilă contradicţia ivită. Potrivit datelor din tabelul 70, aparent, ne-am ocupat de redistribuirea forţei de muncă pe întreprinderi, în realitate, însă, redistribuirea se referă la *structura producţiei în raport cu consumul specific de timp de muncă din perioada de bază* (t_0). Aceasta o arată formulele folosite,iar logica formulelor matematice nu relevă nici un fel de redistribuire, în dinamică, a cheltuielilor de timp de muncă.

Dat fiind că indicele deplasărilor structurale de forma (XI, 5.4a) este un indice derivat, este evident că viciul construcţiei lui este tot derivat, şi de aceea trebuie căutat în formulele celorlalţi doi indici: indicele cu structură variabilă şi indicele cu structură fixă.

Realitatea primului indice nu poate fi pusă la îndoială dat fiind că a numărătorul şi numitorul lui se găsesc mărimi reale. Cît priveşte însă indicele cu structură fixă, în construcţia lui apare o *convenţionalitate* pe care trebuie s-o interpretăm cu toată rigurozitatea pentru a nu da naştere la erori.

Într-adevăr, la numărătorul formulei:

$$I_{\bar{w}}^{\text{str. fixă}} = \frac{\Sigma \cdot \left(\frac{Q_1}{T_1} : \frac{Q_0}{T_0}\right) T_1}{\Sigma T_1} = \frac{\Sigma \frac{T_0}{Q_0} Q_1}{\Sigma T_1} = \frac{\Sigma t_0 Q_1}{\Sigma T_1}$$

avem o mărime convenţională $\Sigma t_0 Q_1$ care reprezintă cheltuielile de timp de muncă pentru producţia din perioada curentă, presupunînd că normele consumului specific de timp de muncă au rămas la nivelul lor din perioada de bază. Caracterul convenţional al acestei mărimi sugerează ideea că tocmai în ea trebuie să căutăm originea contradicţiei semnalate.

Am văzut că această mărime s-a obţinut prin ponderarea indicilor individuali ai productivităţii muncii cu indicatorii de volum ai cheltuielilor de timp de muncă din perioada curentă (T_1). Or, din punct de vedere teoretic, un asemenea sistem de ponderare se dovedeşte a nu fi eficient pentru scopul urmărit, acela de a caracteriza *redistribuirea resurselor de muncă* între diferite ramuri, în perioada curentă. Pentru a atinge acest scop este absolut necesar ca *la construirea indicelui agregat, indicii indi-*

viduali ai productivității muncii (i_w) să fie ponderați cu producția globală din perioada de bază (Q_0).

Într-adevăr, trebuie folosit indicele cu structură fixă de forma :

$$I_{\bar{w}}^{\text{str. fixă}} = \frac{\Sigma\left(\frac{Q_1}{T_1} : \frac{Q_0}{T_0}\right)Q_0}{\Sigma Q_0} = \frac{\Sigma \frac{Q_1 T_0}{T_1}}{\Sigma Q_0} = \frac{\Sigma w_1 T_0}{\Sigma w_0 T_0}. \qquad \text{(XV, 5.5)}$$

Împărțind indicele cu structură variabilă la acest indice cu structură fixă, obținem indicele deplasărilor structurale de altă formă și anume :

$$I_{\bar{w}}^{\text{schimb. str.}} = \frac{\Sigma Q_1 \cdot \Sigma T_0}{\Sigma Q_0 \cdot \Sigma T_1} : \frac{\Sigma \frac{Q_1 T_0}{T_1}}{\Sigma Q_0} = \frac{\Sigma Q_1}{\Sigma T_1} : \frac{\Sigma \frac{Q_0}{T_1} T_1}{\Sigma T_0} = \frac{\Sigma \frac{Q_1}{T_1} T_1}{\Sigma T_1} : \frac{\Sigma \frac{Q_1}{T_1} \cdot T_0}{\Sigma T_0}.$$

Introducînd notația adoptată pentru productivitatea muncii (w_b) și de asemenea, indicatorii de structură $T_0^* = \frac{T_6}{\Sigma T_0}$ și $T_1^0 = \frac{T_1}{\Sigma T_1}$, putem scrie

$$I_{\bar{w}}^{\text{schimb. str.}} = \frac{\Sigma \frac{Q_1}{T_1} T_1}{\Sigma T_1} : \frac{\Sigma \frac{Q_1}{T_1} T_0}{\Sigma T_0} = \frac{\Sigma w_{b_1} T_1}{\Sigma T_1} : \frac{\Sigma w_{b_1} T_0}{\Sigma T_0} = \frac{\Sigma w_{b_1} T_1^*}{\Sigma w_{b_1} T_0^*}. \qquad \text{(XV, 5.6)}$$

După cum se vede, prin construcția acestui indice am obținut elementele ce ne sînt necesare pentru a analiza productivitatea muncii : structura T_1^* se compară cu structura T_0^* și aceasta în raport cu nivelul caracteristicii din perioada curentă.

Dacă aplicăm schema de calcul dată de formula (XV, 5.6) la exemplul din tabelul 69 obținem

$$I_{\bar{w}}^{\text{str. fixă}} = \frac{(1{,}267 \cdot 37\,200) + (1{,}364 \cdot 32\,400) + (1{,}355 \cdot 50\,400)}{37\,200 + 32\,400 + 50\,400} =$$

$$= \frac{47\,132{,}4 + 44\,193{,}6 + 68\,292}{120\,000} = \frac{159\,618}{120\,000} = 1{,}3301 \text{ sau } 133{,}01\%.$$

Folosind indicele cu structură variabilă obținut la p. 403, $I_{\bar{w}} = 1{,}3157$ și împărțindu-l la indicele cu structură fixă obținut acum prin aplicarea formulei (XV, 5.6) avem, în sfîrșit, indicele deplasărilor structurale care ne este necesar pentru analiză :

$$I_{\bar{w}}^{\text{struct.}} = 1{,}3157 : 1{,}3301 = 0{,}9891 \text{ sau } 98{,}91\%.$$

Scanned with OKEN Scanner

Acelaşi rezultat se obţine aplicînd şi direct formula (XV, 5.6) :

$$I_{\bar{w}}^{\text{struct.}} = \frac{\Sigma w_{b_1} T_1^*}{\Sigma w_{b_0} T_0^*} = \frac{(41{,}35 \cdot 0{,}42) + (55{,}63 \cdot 0{,}15) + (66{,}98 \cdot 0{,}33)}{(41{,}35 \cdot 0{,}38) + (55{,}63 \cdot 0{,}28) + (66{,}98 \cdot 0{,}34)} =$$

$$= \frac{52{,}6279}{53{,}2226} = 0{,}9888 \simeq 0{,}989 \text{ sau } 98{,}9\%.$$

Considerînd această mărime şi numai aceasta, putem într-adevăr trage concluzia că în perioada curentă a avut loc o deplasare nefavorabilă a structurii ; se observă că la întreprinderea cu un nivel scăzut al productivităţii muncii a crescut considerabil greutatea specifică a cheltuielilor de timp de muncă (de la 38% la 42%).

Pentru exemplul din tabelul 67 indicele deplasărilor structurale se obţine aplicînd aceeaşi schemă de calcul :

$$I_{\bar{w}}^{\text{str. fixă}} = \frac{(2{,}40 \cdot 37\,200) + (1{,}02 \cdot 32\,400) + (0{,}80 \cdot 50\,400)}{37\,200 + 32\,400 + 50\,400} =$$

$$= \frac{89\,280 + 33\,048 + 40\,320}{120\,000} = \frac{162\,648}{120\,000} = 1{,}3554 \text{ sau } 135{,}54\%,$$

$$I_{\bar{w}}^{\text{schimb. str.}} = 1{,}19 : 1{,}3554 = 0{,}8779. \qquad \text{(XV,5.7)}$$

După cum se vede, de data aceasta indicele deplasărilor structurale caracterizează o situaţie nefavorabilă ($0{,}8779 < 1$) pentru creşterea generală a productivităţii muncii, în comparaţie cu rezultatul obţinut la (XV.3,4 a). Este deci lichidată contradicţia în care ne aflăm pentru că, într-adevăr, în perioada curentă, greutatea specifică a întreprinderilor cu nivelul productivităţii muncii scăzut (nr. 2 şi 3 tabelul 69) a crescut considerabil.

Desigur că orice cercetător poate să rămînă surprins constatînd că şi în formula (XV, 5.6) avem la numitor o mărime convenţională care de data aceasta, însă, ne-a fost de real folos pentru a atinge scopul analitic pe care ni l-am propus. De aceea, se impune o precizare şi anume : viciul ascuns al unor concluzii aparent bune constă nu în faptul însuşi al folosirii unor mărimi convenţionale (cum este de exemplu şi mărimea — $\Sigma w_{b_1} T_0$), ci folosirea lor fără discernămînt (cum a fost, cazul folosirii mărimii $\Sigma t_0 Q_1$ la calculul indicelui cu structură fixă şi apoi a acestuia din urmă la analiza indicelui cu structură variabilă) cu încălcarea principiilor ştiinţifice ale unei analize social-economice concrete.

Am văzut, deci, că sistemul de ponderare folosit în mod obişnuit nu satisface toate cerinţele şi, de aceea, el nu trebuie ridicat la rangul unui principiu general, ci trebuie interpretat raţional în fiecare caz concret de analiză şi aplicat sau nu, în mod diferenţiat, în funcţie de sarcinile concrete ce ni se pun în procesul complex al analizei.

2° *Analiza indicelui preţului de cost cu structură variabilă.* Vom folosi din nou exemplul din tabelul 64 prezentîndu-l de data aceasta sub o altă formă, combinată cu tabelul 65 pentru uşurarea calculelor (tabelul 71).

Scanned with OKEN Scanner

Tabelul 71

Nr. întreprinderii	Felul produsului	Prețul de cost unitar (lei)		Producția în mii unități		Costul total al producției (real și convențional)			
		Perioada de bază	Perioada curentă	Perioada de bază	Perioada curentă	Real	Convențional	Convențional	Real
		z_0	z_1	q_0	q_1	z_0q_0	z_1q_0	z_0q_1	z_1q_1
a	*b*	1	2	3	4	5	6	7	8
1	*A*	25	20	100	600	2 500	2 000	15 000	12 000
	B	35	25	60	400	2 100	1 500	14 000	10 000
						4 600	3 500	29 000	22 000
2	*A*	50	50	400	500	20 000	20 000	25 000	25 000
	B	85	65	500	725	42 500	32 500	61 625	47 125
						62 500	52 500	86 625	72 125
		$\bar{z}_0$	$\bar{z}_1$	$Q_0=\Sigma' q_0$	$Q_1=\Sigma' q_1$	$\bar{z}_0Q_0=$ $=\Sigma z_0q_0$	$\bar{z}_1Q_0$	$\bar{z}_0Q_1$	$\bar{z}_1Q_1=$ $=\Sigma' z_1q_1$
Trust	*A*	45	33,63	500	1 100	22 500	16 815	49 500	37 000
	B	79,64	50,77	560	1 125	44 600	28 431	89 595	57 125
Total general		—	—	—	—	67 100	45 246	139 095	94 125

În acest tabel, datele se referă la două produse *A* și *B*, fabricate la două întreprinderi ale unui trust; în partea de jos a tabelului datele sînt centralizate pe ansamblul trustului, $\bar{z}_1$ și $\bar{z}_0$ fiind valori medii.

Pe baza lor pot fi calculați indicele cu structura fixă și indicele deplasărilor structurale cu scopul de a analiza indicele cu structură variabilă, pe care l-am calculat încă la punctul 2; vezi rezultatul de la (XV, 4.7). În acest scop, calculele au fost incluse în tabelul 72.

Tabelul 72

Produse	Indicii cu structură variabilă			Indicii cu structură fixă			Indicii deplasărilor structurale	Caracterul indicelui (favorabil sau nefavorabil)
	$\bar{z}_1Q_1$	$\bar{z}_0Q_1$	$I_{\bar{z}}$	$\Sigma' z_1q_1$	$\Sigma' z_0q_1$	$I_z(q_1)$	$\frac{I_{\bar{z}}}{I_z(q_1)}=I_q^*(z_0)$	
	1	2	3	4	5	6	7	8
A	37 000	49 500	0,747	37 000	40 000	0,925	0,8075	favorabil
B	57 125	89 595	0,637	57 125	75 625	0,755	0,8437	favorabil
Total	94 125	139 095	0,676	94 125	115 625	0,814 *)	0,8304	favorabil

Se vede că indicele prețului de cost cu structură fixă (coloana 6) și indicele deplasărilor structurale (coloana 7) s-au calculat după metoda recomandată în unele lucrări teoretice și acceptată de obicei în practica statistică curentă.

*) Vezi același rezultat obținut anterior (XV, 2.6).

Scanned with OKEN Scanner

Ce concluzii se pot trage pe baza indicilor care caracterizează deplasările structurale?

Atît la produsul A cît şi la produsul B deplasările structurale au fost favorabile *) reducerii preţului de cost, respectiv cu 19,25% şi 15,63% iar pe ansamblul trustului, pe ambele produse, cu 16,96%.

Trebuie să subliniem însă din nou că aceşti indicatori au înţeles şi trebuie interpretaţi numai în raport cu nivelul preţului de cost din perioada de bază (z_0). Dacă indicatorii de structură q_1^* corespund mai favorabil mărimilor z_0 decît indicatorii q_0^*, aceasta nu înseamnă că ei corespund tot atît de favorabil şi mărimilor z_1. Ca să lămurim complet această problemă am întocmit încă un tabel (tabelul 73) în care sînt ilustrate schimbările structurale şi totodată indicii respectivi au fost calculaţi în raport cu nivelul preţului de cost din perioada curentă (z_1).

Tabelul 73

Produse	Întreprinderi	Producţia în mii unităţi		Indicatorii de structură %		Preţul de cost unitar		Caracterul deplasării structurale	$\frac{I_z}{I_z(q_0)} = I_q^*(z_1)$
		q_0	q_1	$100\,q_0^*$	$100\,q_1^*$	z_0	z_1		
a	*b*	1	2	3	4	5	6	7	8
A	1	100	600	20	54,54	25	20	favorabil	0,764
	2	400	500	80	45,46	50	50		
Total	. . .	500	1 100	100	100	45	23,5454		
B	1	60	400	10,7	35,5	35	25	favorabil	0,836
	2	500	725	89,3	64,5	85	65		
Total	. . .	560	1 125	100,0	100,0	79,64	50,6888		

Este interesant de observat că în ceea ce priveşte produsul A, deplasările structurale au fost evident favorabile şi în raport cu nivelul preţului de cost din perioada curentă; reducerea preţului de cost obţinută pe seama acestui factor este de 23,6%. Într-adevăr, se observă că la întreprinderea nr. 1 în care nivelul preţului de cost în perioada curentă este mai mic (20) decît la întreprinderea nr. 2 (50), greutatea specifică a producţiei reprezintă 54,54% (faţă de 20% din perioada de bază) şi la produsul B au avut loc deplasări structurale favorabile în raport cu z_1; reducerea preţului de cost obţinută pe seama acestui factor este de 16,4 %, deşi greutatea specifică a întreprinderii nr. 2 (cu un preţ de cost mai ridicat) în totalul producţiei acestui produs se menţine mai mare şi în perioada curentă.

*) În cazul dinamicii preţului de cost inegalitatea

$$I_{\text{depl. str.}} < 1$$

ne arată cazul favorabil iar inegalitatea :

$$I_{\text{depl. str.}} > 1,$$

cazul nefavorabil.

Acest fenomen se explică prin faptul că reducerea nivelului relativ al prețului de cost la această întreprindere este de două ori mai mare decît la prima întreprindere.

Prin urmare, o simplă apreciere și caracterizare directă a deplasărilor structurale pe baza datelor din tabelul 72 nu vine în contradicție cu indicatorii deplasărilor structurale calculate în tabelul 73, însă, evident, preferința trebuie acordată indicilor din ultimul tabel. Este firească această preferință prin faptul că practic ne interesează în ce sens anume a influențat fiecare deplasare structurală asupra *nivelului mediu al prețului de cost din perioada curentă*. Or, indicele deplasărilor structurale, calculat prin metoda obișnuită ne dă răspuns la o ipoteză pur teoretică și anume *cum s-ar fi schimbat nivelul mediu al prețului de cost în perioada de bază dacă în această perioadă ar fi existat structura producției din perioada curentă*.

În ultima coloană (8) a tabelului 73 sînt calculați indicii deplasărilor structurale pe baza schemei teoretice pe care o considerăm mai corectă:

— indicele prețului de cost (z) cu structură fixă s-a calculat pe baza ponderilor din perioada de bază (q_0);

— indicatorii de structură q_1^* și q_0^* s-au ponderat cu nivelurile prețului de cost din perioada curentă z_1.

În forma ei generală, această schemă o putem scrie:

$$I_{\bar{z}} = \frac{\bar{z}_1}{\bar{z}_0} = \frac{\Sigma' z_1 q_1}{\Sigma' q_1} : \frac{\Sigma' z_0 q_0}{\Sigma q_0} = \frac{\Sigma' z_1 q_1^*}{\Sigma' z_0 q_0^*} = \frac{\Sigma' z_1 q_0^*}{\Sigma' z_0 q_0^*} \frac{\Sigma' z_1 q_1^*}{\Sigma' z_1 q_0^*} =$$

$$= \frac{\Sigma' z_1 q_0}{\Sigma' z_0 q_0} \cdot \frac{\Sigma z_1 q_1^*}{\Sigma z_1 q_0^*} = I_{z(q_0)} \cdot I_{q^*(z_1)}.$$

Indicii deplasărilor structurale din coloana 8 confirmă aprecierea caracterului deplasărilor structurale dată în coloana 7.

3° *Considerații generale asupra indicelui cu structură variabilă și a elementelor lui componente.* a) *Indicele deplasărilor structurale.* Pentru o parte oarecare a obiectului studiat, indicele individual al indicatorului de structură i_{ψ^*} poate fi scris sub următoarea formă:

$$i_{\psi^*} = \frac{\psi_1^*}{\psi_0^*} = \frac{\psi_1}{\Sigma' \psi_1} : \frac{\psi_0}{\Sigma' \psi_0} = \frac{\psi_1}{\psi_0} : \frac{\Sigma' \psi_1}{\Sigma' \psi_0}$$

care ne arată creșterea relativă a cotei părți date ψ.

În acest caz, indicele agregat al deplasărilor structurale în forma lui obișnuită, cu ponderile α_0, poate fi scris astfel:

$$I_{\psi^*(\alpha_\psi)} = \frac{\Sigma' \alpha_0 \psi_1^*}{\Sigma' \alpha_0 \psi_0^*} = \frac{\Sigma' i_{\psi^*} \alpha_0 \psi_0^*}{\Sigma' \alpha_0 \psi_0^*} = \frac{\Sigma' i_{\psi^*} \alpha_0 \psi_0}{\Sigma' \alpha_0 \psi_0} = \frac{\Sigma' i_{\psi^*} \omega_0}{\Sigma' \omega_0} = \frac{\Sigma' (i_\psi / I_\psi) \omega_0}{\Sigma' \omega_0} \quad (XV, 5.8)$$

în care $\omega_0 = \alpha_0 \psi_0$, iar I_ψ, mărimea medie a indicatorilor de structură.

Se vede că acest indice este o medie aritmetică ponderată a rapoartelor $\frac{i_\psi}{I_\psi}$, ponderile fiind indicatorii caracteristicii funcționale la nivelul lor din perioada de bază.

După părerea lui V. Peregudov, în forma sa (XV, 5.8) indicele deplasărilor structurale poate fi interpretat drept o *dispersie* de un fel deosebit, măsurînd variația indicilor individuali de structură i_ψ *.

Prin analogie cu formula cunoscută a dispersiei (σ^2), caracterizînd variația în cadrul unor distribuții statistice, în speță, mărimile individuale ale indicilor de structură i_ψ sînt comparate cu media lor I_ψ în conformitate cu concepția generală care stă la baza formei matematice a metodei indicilor și anume prin considerarea rapoartelor $\frac{i_\psi}{I_\psi}$. Aceste rapoarte fiind pozitive, nu mai este nevoie să fie ridicate le pătrat — așa cum se procedează la calculul dispersiei. Media lor ponderată ne dă indicele deplasărilor structurale.

Rezultă deci că, prin analogie cu indicatorul dispersiei, indicele deplasărilor structurale măsoară variabilitatea indicilor i_ψ în formă relativă. Cu alte cuvinte, acest indice măsoară *neregularitatea* mișcării elementului de volum ψ pe diferitele părți ale obiectului studiat.

Dacă construim indicele deplasărilor structurale în cealaltă formă a lui, ponderile fiind α_1, ajungem la același indicator al dispersiei relative, sub forma unei medii armonice :

$$I_{\psi^*(\alpha_1)} = \frac{\Sigma'\alpha_1\psi_1^*}{\Sigma\alpha_1\psi_0^*} = \frac{\Sigma'\alpha_1\psi_1^*}{\Sigma'\alpha_1\psi_1^*/i_\psi} = \frac{\Sigma'\alpha_1\psi_1}{\Sigma'\alpha_1\psi_1/i_\psi} = \frac{\Sigma'\omega_1}{\Sigma'\omega_1/i_\psi} = \frac{\Sigma'\omega_1}{\Sigma'\omega_1/(i_\psi/I_\psi)}. \qquad \text{(XV, 5.9)}$$

Dacă toți indicii individuali i_ψ sînt egali între ei și în consecință sînt egali cu I_ψ, indicele deplasărilor structurali este egal cu 1.

b) *Indicele cu structură fixă.* În esență, acest indice este menit să înlăture variabilitatea structurii și să oglindească dinamica relativă a indicatorului calitativ α în forma ei pură. Se pune de fapt problema de a opera o trecere de la *mișcarea reală neregulată* a elementului ψ pe diferite părți ale obiectului studiat, la *mișcarea ipotetică regulată* a aceluiași element. Mișcarea reală generală a elementului ψ este caracterizată prin indicele general I_ψ și tocmai acest ritm este menținut și în cadrul mișcării ipotetice regulate. Prin urmare, de la mișcarea reală a elementului ψ caracterizată prin mărimi diferite ale indicilor individuali i_ψ are loc trecerea spre mișcarea ipotetică, ajustată, a elementului ψ în care evident, toate mărimile i_ψ sînt egale între ele și totodată egale cu indicele general I_ψ ($i_\psi = I_\psi = c$ = constantă). Aceasta înseamnă că noi presupunem o asemenea mișcare regulată a elementului ψ în care are loc trecerea de la mărimile reale ψ_0 la anumite mărimi convenționale $\psi_1' = c\,\psi_0$, păstrîndu-se, cu acest prilej, creșterea generală a elementului de volum de la $\Sigma'\psi_0$ la $\Sigma'\psi_1'$ și eliminîndu-se neregularitatea mișcării reale de la ψ_0 la ψ_1.

Rezultă că se poate construi un indice al caracteristicii calitative α pe baza ponderilor ψ_1' care, evident, va fi egal cu indicele construit pe baza ponderilor ψ_0

$$I_{\alpha(\psi_1')} = \frac{\Sigma'\alpha_1\psi_1'}{\Sigma'\alpha_0\psi_1'} = \frac{\Sigma'\alpha_1 c\cdot\psi_0}{\Sigma'\alpha_0 c\cdot\psi_0} = \frac{\Sigma'\alpha_1\psi_0}{\Sigma'\alpha_0\psi_0} = I_{\alpha(\psi_0)}.$$

Scanned with OKEN Scanner

Așadar, mișcarea reală a elementului ψ de la nivelul de bază ψ_0 la cel din perioada curentă ψ_1 are loc în două etape care, schematic, pot fi reprezentate astfel :

$$\psi_0 \rightarrow \psi_1' \rightarrow \psi_1 .$$

Prima etapă corespunde mișcării fără deplasări structurale, a doua mișcării influențate și de deplasările structurale.

Grafic, această mișcare poate fi reprezentată ca în figura 39.

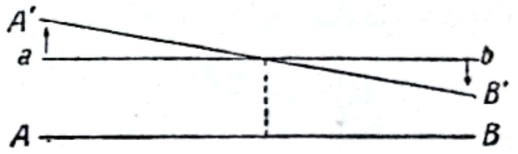


Fig. 39

Dacă o dreaptă AB, ca rezultat al mișcării de translație și de rotație care au avut loc simultan, a ajuns în poziția $A'B'$ atunci, în scopuri analitice, noi ne putem reprezenta că această mișcare se poate descompune în mișcarea de translație (în poziția ab) și mișcarea de rotație (trecerea de la poziția ab la poziția $A'B'$) *).

Prin urmare, descompunerea indicelui cu structură variabilă în doi indici

$$I_{\bar{\alpha}} = \frac{\bar{\alpha}_1}{\bar{\alpha}_0} = \frac{\Sigma' \alpha_1 \psi_0}{\Sigma' \alpha_0 \psi_0} \cdot \frac{\Sigma' \alpha_1 \psi_1^*}{\Sigma' \alpha_0 \psi_0^*} = I_{\alpha(\psi_0^*)} \; I_{\psi^*(\alpha_1)}$$

în ipoteza egalității

$$(\psi_1')^* = (c\psi_0)^* = \psi_0^*$$

poate fi prezentată astfel :

$$I_{\bar{\alpha}} = \frac{\Sigma' \alpha_1 \psi_0^*}{\Sigma' \alpha_0 \psi_0^*} \cdot \frac{\Sigma' \alpha_1 \psi_1^*}{\Sigma' \alpha_1 \psi_0^*} = \frac{\Sigma' \alpha_1 (\psi_1')^*}{\Sigma' \alpha_0 (\psi_1')^*} \cdot \frac{\Sigma' \alpha_1 \psi_1^*}{\Sigma' \alpha_1 (\psi_1')^*} = \frac{\Sigma' \alpha_1 \psi_1'}{\Sigma' \alpha_0 \psi_1'} \cdot \frac{\Sigma' \alpha_1 \psi_1^*}{\Sigma' \alpha_1 (\psi_1')^*} .$$

Să aplicăm această schemă teoretică la exemplul nostru din tabelul 74 alegînd pentru analiză produsul B.

Tabelul 74

Indicatori	Întreprinderea 1	Întreprinderea 2	Total
Volumul producției din perioada de bază (q_0)	60	500	560
Structura producției în perioada de bază (q_0^*)	0,107	0,893	1,00
Volumul convențional al producției din perioada curentă cu structura producției din perioada de bază (q_1')	120,375	1 004,625	1 125
Volumul producției din perioada curentă (q_1)	400	725	1 125
Structura producției din perioada curentă (q_1^*)	0,355	0,645	1,00
Prețul de cost în perioada de bază (z_0)	35	85	79,64
Prețul de cost în perioada curentă (z_1)	25	65	50,77

*) Procedeul de reprezentare folosit de V. Peregudov este cunoscut în știință ca procedeu de analiză a fenomenelor complexe în dinamică.

Mișcarea volumului total al producției din perioada de bază $\Sigma' q_0 = 560$ către volumul ei din perioada curentă $\Sigma' q_1 = 1\,125$ are loc, după cum știm, în două etape:

— În prima etapă are loc mișcarea de translație, *fără schimbări structurale*, obținîndu-se indicatorii q_1'. Dinamica prețului de cost al produsului B (tabelul 74), pe ansamblul trustului în această etapă, va fi caracterizată prin următorul indice:

$$\frac{\Sigma' z_1 q_1'}{\Sigma' z_0 q_1'} = \frac{(25.120{,}375) + (65.1004{,}625)}{(35.120{,}375) + (85{,}1004{,}625)} = \frac{68\,310}{89\,606{,}250} \cong 0{,}762.$$

Același indice se obține și pe baza ponderilor q_0:

$$\frac{\Sigma z_1 q_0}{\Sigma z_0 q_0} = \frac{(25.60) + (65.500)}{(35.60) + (85.500)} = \frac{34\,000}{44\,600} \cong 0{,}762.$$

— În a doua etapă au loc deplasările structurale, fără *mișcarea de translație*, cu alte cuvinte are loc trecerea de la mărimile convenționale q_1' la mărimile reale q_1.

Schimbarea prețului de cost în această etapă este caracterizată prin indicile:

$$\frac{\Sigma z_1 q_1}{\Sigma z_1 q_1'} = \frac{(25.400) + (65.725)}{(25.120{,}375) + (65.1004{,}625)} = \frac{57\,125}{68\,310} = 0{,}836.$$

Evident că același rezultat se poate obține și cu ajutorul indicelui deplasărilor structurale

$$\frac{\Sigma z_1 q_1'}{\Sigma z_1 q_0'} = \frac{(25.0{,}355) + (65.0{,}645)}{(25.0{,}107) + (65.0{,}893)} = \frac{50{,}800}{60{,}720} = 0{,}836.$$

Riguros vorbind, indicele cu structură variabilă al prețului de cost al produsului B egal cu 0,637 se poate descompune în felul următor:

$$I_{z_B} = \frac{\bar{z}_{1B}}{\bar{z}_{0B}} = \frac{\Sigma z_1 q_0'}{\Sigma z_0 q_0'} \cdot \frac{\Sigma z_1 q_1'}{\Sigma z_1 q_0'} = \frac{60{,}720}{79{,}650} \cdot \frac{50{,}800}{60{,}720} = \quad \text{(XV, 5.10)}$$

$$= 0{,}762 \cdot 0{,}836 = 0{,}637.$$

Formula (XV, 5.10) poate fi folosită și la analiza sporului absolut (pozitiv sau negativ) al prețului de cost mediu:

$$\bar{z}_1 - \bar{z}_0 = \Sigma z_1 q_1' - \Sigma z_0 q_0' = (\Sigma z_1 q_0' - \Sigma z_0 q_0') + (\Sigma z_1 q_1' - \Sigma z_1 q_0') =$$

$$= (60{,}720 - 79{,}650) + (50{,}800 - 60{,}720) = -18{,}930 - 9{,}920 = -28{,}850.$$

Dacă revenim acum asupra schemei teoretice folosită în mod obișnuit la analiza indicelui cu structură variabilă în cazul produsului B, avem

$$I_{z_B} = \frac{\bar{z}_{1B}}{\bar{z}_{0B}} = \frac{\Sigma z_1 q_1}{\Sigma z_0 q_1} \cdot \frac{\Sigma z_0 q_1'}{\Sigma z_0 q_0'} = 0{,}755 \cdot 0{,}8437 = 0{,}637.$$

Scanned with OKEN Scanner

Devine clar acum că o asemenea descompunere a indicelui cu structură variabilă trebuie să reprezinte mișcarea inversă a elementului de volum al fenomenului complex, după schema $q_0 \leftarrow q_0' \leftarrow q_1'$. Aceasta înseamnă, evident, că întreaga analiză este orientată spre nivelul prețului de cost din perioada de bază, în ipoteza — lipsită de interes practic — că structura producției ar fi cea din perioada curentă.

6. Analiza indicelui caracteristicii funcționale în ipoteza unei producții eterogene

Prin conținutul lor, indicii de tipul A (XV, 3.3) sînt și ei tot atît de complecși ca indicii cu structură variabilă. Într-adevăr, și acești indici exprimă, sintetic, două dinamici interdependente: dinamica relativă a caracteristicii calitative (α) și dinamica caracteristicii de volum (ψ), iar aceasta din urmă sub forma a două mișcări — simultane — mișcarea de translație (fără deplasări structurale) și mișcarea internă, de regrupare, a elementelor de volum (dinamica schimbărilor structurale).

Se pune, de aceea, în mod firesc problema ca în procesul analizei unui indice al caracteristicii funcționale să se determine un indice al deplasărilor structurale *chiar și în ipoteza cînd elementul* ψ *nu este însumabil*, pentru că în realitate, deplasări structurale au loc și în acest caz.

Un asemenea indice a fost propus tot de V. Peregudov *). Vom prezenta aici schema teoretică a acestui indice, ea urmînd a-și găsi mai tîrziu o interesantă aplicație la studiul indicilor teritoriali (vezi capitolul următor).

Fie $i_\psi', i_\psi'', i_\psi''', \ldots, i_\psi^n$, n indici individuali ai diferitelor produse, fiecare cu unitatea lui specifică de măsură și $I_{\psi(\alpha)}$, indicele lor mediu sau — ceea ce este același lucru — indicele agregat care caracterizează ritmul mediu al mișcării de translație.

Problema se pune de a găsi un indice al deplasărilor structurale, element component al indicelui fenomenului complex $\omega = \alpha\psi$.

Pentru rezolvarea ei trebuie să facem unele precizări prealabile pe marginea a două situații diferite:

a) cînd elementul ψ este însumabil (producția este omogenă) și
b) cînd elementul ψ nu este însumabil (producția este eterogenă).

Descompunerea indicelui fenomenului complex capătă în mod corespunzător și ea două forme diferite:

a) $$I_{\alpha\psi} = \frac{\Sigma\alpha_1\psi_1}{\Sigma\alpha_0\psi_0} = \frac{\Sigma\alpha_1\psi_1'}{\Sigma\alpha_0\psi_0'} \cdot \frac{\Sigma\psi_1}{\Sigma\psi_0} = \frac{\Sigma\alpha_1\psi_0}{\Sigma\alpha_0\psi_0} \cdot \frac{\Sigma\alpha_1\psi_1'}{\Sigma\alpha_1\psi_0'} \cdot \frac{\Sigma\psi_1}{\Sigma\psi_0};$$

b) $$I_{\alpha\psi} = \frac{\Sigma\alpha_1\psi_1}{\Sigma\alpha_0\psi_0} = \frac{\Sigma\alpha_1\psi_1}{\Sigma\alpha_0\psi_1} \cdot \frac{\Sigma\alpha_0\psi_1}{\Sigma\alpha_0\psi_0}.$$

*) V. Peregudov, *Ucionlie zapiski po statistike*, vol. I, Moskva, 1955.

Scanned with OKEN Scanner

Se vede că în situaţia a) (producţia omogenă), indicele $I_{\alpha\psi}$ se descompune *în trei părţi* distincte: indici extremi reprezintă mişcarea de translaţie a celor două elemente componente α şi ψ, iar cel de la mijloc, mişcarea de regrupare internă (dinamica schimbărilor structurale).

În situaţia b), acelaşi indice se descompune numai în *două părţi* componente, fiecare reprezentînd, respectiv, dinamica elementului α şi elementului ψ — în mişcarea lor de translaţie; în această situaţie nu se vede nici o măsură statistică a mişcării de regrupare internă deşi o asemenea mişcare are loc în realitate. Relevăm deci o contradicţie între forma descompusă a indicelui $I_{\alpha\psi}$ în cazul unei producţii eterogene şi conţinutul real al procesului economic.

Este evident însă că şi în situaţia b), dacă vrem să relevăm numai mişcarea de translaţie a elementului ψ, trebuie să ponderăm elementul α cu ψ_0. De aceea, se impune ca din indicele $I_{\alpha(\psi_1)}$ să izolăm indicele elementului α în condiţiile inexistenţei regrupărilor interne ale elementului ψ; cu alte cuvinte să scoatem în evidenţă indicele $I_{\alpha(\psi_0)}$ pe de o parte, şi un coeficient $k_{\alpha\psi}$ pe de altă parte.

Vom avea deci:

$$\frac{I_{\alpha(\psi_1)}}{I_{\alpha(\psi_0)}} = k_{\alpha\psi}$$

sau

$$k_{\alpha\psi} = \frac{\Sigma\alpha_1\psi_1}{\Sigma\alpha_0\psi_1} : \frac{\Sigma\alpha_1\psi_0}{\Sigma\alpha_0\psi_0} = \frac{\Sigma\alpha_1\psi_1 \cdot \Sigma\alpha_0\psi_0}{\Sigma\alpha_0\psi_1 \cdot \Sigma\alpha_1\psi_0}.$$

Pentru elementul ψ putem deduce, în mod analog, aceeaşi mărime.

$$k_{\alpha\psi} = \frac{\Sigma\alpha_1\psi_1}{\Sigma\alpha_1\psi_0} : \frac{\Sigma\alpha_0\psi_1}{\Sigma\alpha_0\psi_0} = \frac{\Sigma\alpha_0\psi_0 \cdot \Sigma\alpha_1\psi_1}{\Sigma\alpha_1\psi_0 \cdot \Sigma\alpha_0\psi_1}$$

Observăm că, prin conţinutul său, coeficientul $k_{\alpha\psi}$ prezintă o simetrie perfectă în ceea ce priveşte elementele α şi ψ.

Acest coeficient nu poate fi atribuit, de aceea, în mod exclusiv nici unuia din aceste elemente, el aparţinînd în egală măsură şi unuia şi altuia:

Se poate face următoarea dublă transformare:

$$k_{\alpha\psi} = \frac{\Sigma\alpha_0\psi_0 \cdot \Sigma\alpha_1\psi_1}{\Sigma\alpha_1\psi_0 \cdot \Sigma\alpha_0\psi_1} = \frac{\Sigma i_\psi(\alpha_1\psi_0)}{\Sigma(\alpha_1\psi_0)} : \frac{\Sigma i_\psi(\alpha_0\psi_0)}{\Sigma(\alpha_0\psi_0)} = \frac{\Sigma i_\psi(\alpha_1\psi_0)^*}{\Sigma i_\psi(\alpha_0\psi_0)^*}$$

$$k_{\alpha\psi} = \frac{\Sigma\alpha_0\psi_0 \cdot \Sigma\alpha_1\psi_1}{\Sigma\alpha_1\psi_0 \cdot \Sigma\alpha_0\psi_1} = \frac{\Sigma i_\alpha(\alpha_0\psi_1)}{\Sigma(\alpha_0\psi_1):} : \frac{\Sigma i_\alpha(\alpha_0\psi_0)}{\Sigma(\alpha_0\psi_0)} = \frac{\Sigma i_\alpha(\alpha_0\psi_1)^*}{\Sigma i_\alpha(\alpha_0\psi_0)^*}$$

În ambele cazuri avem indici ai deplasărilor structurale de un fel deosebit. În primul caz se compară structurile $(\alpha_1\psi_0)^*$ şi $(\alpha_0\psi_0)^*$ iar în al doilea, structurile $(\alpha_0\psi_1)^*$ şi $(\alpha_0\psi_0)^*$. Aceste structuri sînt estimate nu în raport cu un nivel fix al elementului α — aşa cum se procedează la construirea indicelui obişnuit al deplasărilor structurale — ci în raport cu indicatorii dinamicii i_ψ şi i_α care îndeplinesc funcţia unităţilor de comensurare (ponderare).

Scanned with OKEN Scanner

Esența coeficientului $k_{\alpha\psi}$ poate fi scoasă în evidență și mai pregnant dacă îi aducem unele transformări, înlocuind mărimea $\alpha_0\psi_0$ cu ω_0.

Vom avea, astfel :

$$k_{\alpha\psi} = \frac{\Sigma\alpha_0\psi_0 \cdot \Sigma\alpha_1\psi_1}{\Sigma\alpha_1\psi_0 \cdot \Sigma\alpha_0\psi_1} = \frac{\Sigma\omega_0 \cdot \Sigma i_\alpha i_\psi \omega_0}{\Sigma i_\alpha\omega_0 \cdot \Sigma i_\psi\omega_0} = \frac{\Sigma i_\alpha i_\psi\omega_0}{\Sigma\omega_0} : \left(\frac{\Sigma i_\alpha\omega_0}{\Sigma\omega_0} \cdot \frac{\Sigma i_\psi\omega_0}{\Sigma\omega_0}\right) = \frac{\overline{(i_\alpha \cdot i_\psi)}}{\bar{i}_\alpha \cdot \bar{i}_\psi}.$$

Valoarea coeficientului $k_{\alpha\psi}$ se poate exprima cu ajutorul coeficientului de corelație a două variabile i_α și i_ψ.

Dacă pentru două șiruri de date

$$i'_\alpha,\ i''_\alpha,\ \ldots,\ i^n_\alpha$$

și

$$i'_\psi,\ i''_\psi,\ \ldots,\ i^n_\psi$$

cu ponderile corespunzătoare

$$\omega'_0,\ \omega''_0,\ \ldots,\ \omega^n_0$$

s-ar calcula coeficientul de corelație $\rho_{i_\alpha i_\psi}$, atunci el ar avea următoarea formă :

$$\rho_{i_\alpha i_\psi} = \frac{\Sigma\omega_0 \Sigma i_\alpha i_\psi \omega_0 - \Sigma i_\alpha\omega_0 \cdot \Sigma i_\psi\omega_0}{\sigma_{i_\alpha} \cdot \sigma_{i_\psi}(\Sigma\ \omega_0)^2} = \frac{\Sigma\omega_0 \Sigma i_\alpha i_\psi\omega_0}{\sigma_{i_\alpha}\sigma_{i_\psi}(\Sigma\omega_0)^2} - \frac{\Sigma i_\alpha\omega_0 \cdot \Sigma i_\psi\omega_0}{\sigma_{i_\alpha} \cdot \sigma_{i_\psi}(\Sigma\omega_0)^2} = A - B.$$

Rezultă, așadar, că coeficientul de corelație $\rho_{i_\alpha i_\psi}$, care caracterizează intensitatea legăturii reciproce dintre indicii i_α și i_ψ se obține prin compararea a două mărimi A și B, sub forma diferenței $A - B$. Trebuie arătat însă, că în acest caz, coeficientul de corelație nu are nici o legătură cu vreo schemă probabilistică, el reprezintă pur și simplu un indicator statistic al gradului de concordanță a două șiruri de date i_α și i_ψ.

Dacă mărimile A și B vor fi comparate potrivit metodei specifice a indicilor, făcîndu-se raportul $\frac{A}{B}$, atunci obținem tocmai coeficientul $k_{\alpha\psi}$:

$$\frac{A}{B} = \frac{\Sigma\omega_0 \cdot \Sigma i_\alpha i_\psi \omega_0}{\Sigma i_\alpha\omega_0 \Sigma i_\psi\omega_0} = k_{\alpha\psi}.$$

Între coeficientul de corelație $\rho_{i_\alpha i_\psi}$ și $k_{\alpha\psi}$ pot fi stabilite următoarele relații :

dacă $\rho_{i_\alpha i_\psi} < 0$ atunci $A < B$ și $k_{\alpha\psi} < 1$
„ $\rho_{i_\alpha i_\psi} = 0$ „ $A = B$ și $k_{\alpha\psi} = 1$
„ $\rho_{i_\alpha i_\psi} > 0$ „ $A > B$ și $k_{\alpha\psi} > 1$

Prin urmare, coeficientul $k_{\alpha\psi}$ este legat nemijlocit de coeficientul de corelație $\rho_{i_\alpha i_\psi}$. Este adevărat că el nu are limitele ± 1 care marchează trecerea la relații liniare exacte, dar, în condițiile metodei indicilor, acest lucru nu este esențial. În schimb, coeficientul $k_{\alpha\psi}$ caracterizează *legătura dintre elementele α și ψ în procesul dezvoltării lor dinamice* sau cu alte cuvinte *dinamica interdependentă a elementelor α și ψ.*

Scanned with OKEN Scanner

Coeficientul $k_{\alpha\psi}$ este denumit, de V. Peregudov, *indice al legăturii reciproce a elementelor* α *şi* ψ.

Întrucît

$$I_{\alpha(\psi_1)} = k_{\alpha\psi} \cdot I_{\alpha(\psi_0)}$$

şi

$$I_{\psi(\alpha_1)} = k_{\alpha\psi} \cdot I_{\psi(\alpha_0)},$$

cele două variante în care se poate descompune indicele fenomenului complex $I_{\alpha\psi}$,

$$I_{\alpha\psi} = I_{\alpha(\psi_1)} \cdot I_{\psi(\alpha_0)}$$

şi

$$I_{\alpha\psi} = I_{\alpha(\psi_0)} \cdot I_{\psi(\alpha_1)}$$

pot fi înlocuite cu una singură, avînd *trei elemente* componente:

$$I_{\alpha\psi} = k_{\alpha\psi} \cdot I_{\alpha(\psi_0)} \cdot I_{\psi(\alpha_0)}.$$

Schema analitică a dinamicii fenomenului complex $\alpha\ \psi$ nu va fi deci de forma

$$\alpha_0\psi_0 \rightarrow \alpha_0\psi_1 \rightarrow \alpha_1\psi_1$$

în care nu este scoasă în evidenţă corelaţia reală care există între cele două elemente α şi ψ, ci de forma din figura 40.

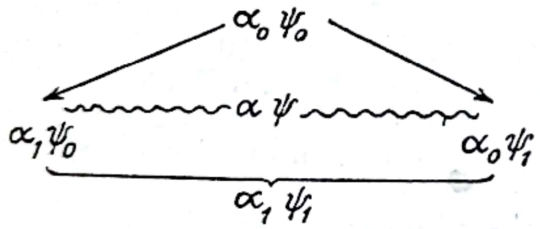


Fig. 40

În această schemă este scos în evidenţă şi factorul legăturii reciproce, reprezentat simbolic prin $(\alpha\psi)$.

Mişcarea, dinamica fenomenului complex, este complexă prin ea însăşi, conţinînd şi legătura interdependentă, în dinamică, a elementelor componente.

În particular, se poate observa că orice indice care se construieşte pe baza ponderilor date de structura din perioada curentă a elementului de volum ψ, cum ar fi, de exemplu, indicele $I_{\alpha(\psi_1)}$, va conţine şi un element care nu-i aparţine exclusiv şi anume, mărimea $k_{\alpha\psi}$.

Scanned with OKEN Scanner

Capitolul XVI

Indici teritoriali

1. Considerații introductive

Categoria statistică a indicilor teritoriali ocupă un loc aparte în cadrul indicilor statici și, în general, în cadrul indicilor ca metodă specifică a statisticii social-economice. În acest capitol vom încerca să relevăm și să sintetizăm problemele teoretice și practice esențiale ale elaborării indicilor teritoriali.

Prin indici teritoriali se înțeleg acei indici prin care se compară datele unei regiuni cu datele unei alte regiuni*) sau chiar datele mai multor regiuni între ele. De pildă, dacă ne punem problema de a compara recolta medie la un ha de culturi cerealiere, obținută în diferite regiuni, cu recolta medie a unei regiuni fruntașe, este cazul să construim mai mulți indici teritoriali; aceiași indici trebuie să-i construim și în cazul cînd ne punem problema de a compara nivelul productivității muncii din diferite întreprinderi ale unei anumite ramuri industriale, cu productivitatea muncii atinsă în întreprinderea cu cel mai înalt nivel de înzestrare tehnică, din cadrul aceleiași ramuri etc.

După cum se vede, în exemplele alese se pune problema de a compara, pe plan spațial, static, cîțiva indicatori centralizatori cu caracter de medii. A realiza o asemenea comparație implică, însă, cunoscutele dificultăți legate de faptul că ea se efectuează în condițiile deosebirilor în structura colectivităților comparate, adeseori deosebit de pronunțate.

Considerăm că rezolvarea acestor dificultăți în domeniul indicilor teritoriali nu se poate face decît pe cale de consecință, rezolvînd, în prealabil, două probleme metodologice de prim rang **) și anume:

— alegerea bazei de comparație și a sistemului de ponderi;

*) Noțiunea de regiune capătă aici un sens foarte larg; poate fi vorba de o țară, de o regiune administrativă, economică, istorică sau chiar de un raion, întreprindere etc.

**) Nu ne vom opri și asupra problemelor legate de alegerea formei indicilor teritoriali, întrucît, după cum se va putea vedea, forma principală folosită la construirea indicilor teritoriali este aceea a indicilor agregați.

— asigurarea corelației de sistem între indicii factoriali și indicele fenomenului complex.

Trebuie să precizăm, însă, că cercetările științifice pentru rezolvarea problemelor metodologice ale teoriei indicilor teritoriali sînt încă în curs, constituind obiectul unor interesante discuții publicate în revista „Vestnik Statistiki". În cadrul lor și în deosebi pe marginea părerilor formulate inițial de S. Iugenburg și L. M. Satunovski, aspecte teoretice și practice noi, interesante, au fost relevate de cunoscuții statisticieni și economiști sovietici G. Baklanov, P. Joffe, L. Kazineț, V. Peregudov și Z. Sidlik în studiile publicate în aceeași revistă spre sfîrșitul anului 1958 și în cursul anului 1959.

Dat fiind că discuția cu privire la metodele de calcul al indicilor teritoriali prezintă un deosebit interes teoretic și practic, ne vom opri pe scurt asupra problemelor noi ridicate de autorii amintiți, făcînd, în încheiere, cîteva constatări și trăgînd unele concluzii generale.

2. Alegerea bazei de comparație (raportare) și a sistemului de ponderi

Alegerea bazei de comparație și a sistemului corespunzător de ponderi la construirea indicilor teritoriali este o problemă ce se rezolvă în funcție de scopul concret și sarcinile analizei ce ne propunem.

Dacă ne referim la cazul cînd se pune problema de a compara nivelul productivității muncii din mai multe întreprinderi industriale cu productivitatea muncii obținută la o întreprindere cu cel mai înalt nivel de înzestrare tehnică din cadrul aceleiași ramuri, este evident că, în acest caz, drept bază de comparație ne vor servi indicatorii acestei din urmă întreprinderi. Vom construi atîția indici teritoriali cîte întreprinderi au fost luate în studiu, la numitorul fiecărui indice figurînd productivitatea muncii de la întreprinderea cu cel mai înalt nivel de înzestrare tehnică.

Alegerea bazei de comparație în acest exemplu a decurs în mod firesc din însuși modul în care s-a pus problema concretă de analiză.

Să presupunem un alt caz. Ni se pune problema de a compara nivelul salariului mediu al muncitorilor din cinci întreprinderi, aparținînd aceleiași ramuri industriale. Astfel pusă, problema nu se mai poate rezolva ca în cazul precedent — fără riscul unei soluții arbitrare — alegîndu-se drept bază de comparație salariul mediu în una din cele cinci întreprinderi. Aici nu avem nici un criteriu obiectiv, fundamentat din punct de vedere economic, pentru a proceda la o asemenea alegere, preferînd o întreprindere sau alta. Este lesne de înțeles că din cauza deosebirilor în structura grupelor de muncitori din cele cinci întreprinderi, alegerea unei întreprinderi sau alteia ca bază de comparație poate să ducă la mărimi ale indicilor teritoriali cu totul contradictorii.

Așadar, alegerea bazei de comparație în cazul problemelor statistice concrete, de tipul acesteia din urmă, ridică, de la început, o mare dificultate.

Scanned with OKEN Scanner

O vom lămuri cu ajutorul unui exemplu (tabelul 75) folosit de autorul sovietic S. Iughenburg *) (date convenționale).

Tabelul 75

Ramuri industriale	Producția globală (în milioane ruble)		Numărul mediu anual al muncitorilor		Numărul de om-ani la 1 milion ruble producție	
	Regiunea A	Regiunea B	Regiunea A	Regiunea B	Regiunea A	Regiunea B
Ramura I	100	30	16 000	6 000	160	200
Ramura II	60	400	2 400	12 000	40	30
Total	160	430	18 400	18 000	115,0	41,9

După cum se vede, sînt luate două regiuni, din care în una (A) sînt dezvoltate puternic ramurile industriei miniere și carbonifere (ramura I), iar în a doua (B), unele ramuri ale industriei ușoare și alimentare cu o producție medie ridicată, în ruble, pe un muncitor (ramura II).

Problema ce se pune este aceea de a compara volumul fizic al producției celor două regiuni, calculată în om-ani. În acest scop, autorul se folosește de următoarea formulă a indicelui volumului fizic al producției ($I_{\text{vol. fiz.}}$) :

$$I_{\text{vol. fiz.}} = \frac{\sum \frac{T_0}{q_0} q_1}{\Sigma T_0},$$ **)

Luînd ca bază de comparație, pe rînd, cheltuielile medii de muncă vie pe unitate de produs $\left(\frac{T_0}{q_0}\right)$, în regiunea A și apoi în regiunea B, autorul a obținut următorii indici teritoriali :

a) Pentru regiunea B (baza de comparație fiind regiunea A) :

$$I^{\text{ter.}}_{\text{vol. fiz.}\,(B)} = \frac{\sum \frac{T_0}{q_0} q_1}{\sum \frac{T_0}{q_0} q_0} = \frac{\sum \frac{T_0}{q_0} q_1}{\Sigma T_0} = \frac{16 \cdot 30 + 40 \cdot 400}{16\,000 + 2\,400} = \frac{20\,800 \text{ om-ani}}{18\,400 \text{ om-ani}}.$$

*) S. Iughenburg, *Nekotorie voprosi iscisleniia teritorialnih indeksov*. Vestnik Statistiki, 4 (1958).

**) Deosebit de interesante sînt argumentele invocate de autor pentru a demonstra teza pe care o formulează, susținînd în esență că dinamica volumului fizic al producției se determină prin schimbarea „cheltuielilor medii de timp de muncă vie, socialmente necesară, față de nivelul productivității muncii din perioada de bază". Aceasta se oglindește și în formula de mai sus a indicelui. Ea derivă din produsul a doi indici factoriali, și anume indicele productivității muncii exprimat în cheltuieli de timp de muncă și indicele cheltuielilor totale de timp de muncă :

$$\frac{\sum \frac{T_0}{q_0} q_1}{\Sigma T_1} \cdot \frac{\Sigma T_1}{\Sigma T_0} = \frac{\sum \frac{T_0}{q_0} q_1}{\Sigma T_0}.$$

Se observă, de altfel, că dacă în loc de ΣT_0 vom pune la numitor $\sum \frac{T_0}{q_0} q_0$, obținem formula cunoscută a indicelui volumului fizic.

Rezultă că volumul producţiei în regiunea B este mai mare decît în regiunea A cu 13%.

b) Pentru regiunea A (baza de comparaţie constituind-o regiunea B) :

$$I^{\text{ter.}}_{\text{vol. fiz.}\,(A)} = \frac{\sum \frac{T_0}{q_0} q_1}{\sum \frac{T_0}{q_0} q_0} = \frac{\sum \frac{T_0}{q_0} q_1}{\Sigma T_0} = \frac{200 \cdot 100 + 30 \cdot 60}{6\,000 + 12\,000} = \frac{21\,800 \text{ om-ani}}{18\,000 \text{ om-ani}}.$$

Compararea producţiei celor două regiuni, în a doua ipoteză, ne dă un rezultat — după cum se vede — cu totul opus primului : a rezultat că producţia regiunii A este cu 21% mai mare decît producţia regiunii B.

Constatînd această contradicţie, S. Iughenburg ajunge la concluzia că ea se datorează deosebirilor mari în structura producţiei din regiunile comparate. Această împrejurare face ca sistemul obişnuit de ponderare, folosit la calculul indicilor dinamicii fenomenelor, să nu poată funcţiona în domeniul indicilor teritoriali. Pentru rezolvarea acestei probleme cercetătorii sovietici au propus diferite metode.

3. Metode de calcul

Metoda lui S. Iughenburg. S. Iughenburg propune folosirea metodei coeficienţilor standard *) care elimină influenţa deosebirilor structurale la compararea în spaţiu, pe plan interregional, a caracteristicilor studiate.

*) Metoda coeficienţilor standard a fost elaborată iniţial în statistica demografică în scopul înlăturării influenţei pe care o exercită deosebirile din structura populaţiei asupra unor indicatori ce se iau în studiu.

Dacă folosim următoarea notaţie :

x_1, coeficienţii de mortalitate pe vîrste ai unei grupe de populaţie ;
l_1, ponderile respective la aceeaşi grupă de populaţie (pe vîrste) ;
x_0, coeficienţii de mortalitate pe vîrste pentru o altă grupă de populaţie ;
l_0, ponderile respective ale celei de-a doua grupe de populaţie (pe vîrste) ;
x, coeficienţii de mortalitate standard ;
l, ponderile standard ;

raportul dintre mortalitatea medie a celor două categorii de populaţie va avea următoarea formă de indice :

$$\frac{\Sigma x_1 l_1}{\Sigma x_0 l_0}.$$

Aplicînd însă metoda coeficienţilor standard, în cele două variante, obţinem indicele coeficientului standard al mortalităţii :

a) metoda directă

$$I_{dir.} = \frac{\Sigma x_1 l}{\Sigma x_0 l};$$

b) metoda indirectă

$$I_{ind.} = \frac{\Sigma x_1 l_1}{\Sigma x l_1} : \frac{\Sigma x_0 l_0}{\Sigma x l_0}.$$

În legătură cu această problemă vezi exemplul aplicării metodei coeficienţilor standard în cursul : *Demograficeskaia statistika*, de prof. A. I. Boiarski şi docent P. P. Suşerin, Moskva, 1951, p. 123—127 ; în acest exemplu se arată două variante ale metodei (directă şi indirectă), cu

Aplicată la indicii teritoriali, metoda coeficienţilor standard exclude sistemul de ponderare a indicilor cu elementele de structură ale uneia din regiunile comparate; ea presupune, în schimb, alegerea unor ponderi care să oglindească structura pe un teritoriu mult mai larg, cuprinzînd toate regiunile comparate, sau chiar o întreagă ţară.

Avantajul acestei metode — subliniază S. Iughenburg — constă în aceea că oferă posibilitatea de a se construi şi pentru indicii teritoriali un sistem de ponderi, pe baze obiective.

Aplicînd-o la exemplul de mai sus, autorul a luat drept ponderi standard cheltuielile medii pe ambele regiuni luate la un loc ($\bar{t}$), adică cheltuielile medii de muncă pe fiecare ramură (om-ani la 1 milion ruble-producţie):

— în ramura industrială I:

$$\bar{t}_{\text{I}} = \frac{\Sigma T_{\text{I}}}{\Sigma q_{\text{I}}} = \frac{16\,000 + 6\,000}{100 + 30} = \frac{22\,000}{130} = 169{,}2 \text{ om-ani};$$

— în ramura industrială II:

$$\bar{t}_{\text{II}} = \frac{\Sigma T_{\text{II}}}{\Sigma q_{\text{II}}} = \frac{2\,400 + 12\,000}{60 + 400} = \frac{14\,400}{460} = 31{,}3 \text{ om-ani}.$$

El a calculat apoi indicatorii volumului producţiei pentru regiunile respective, exprimaţi în om-ani necesari pentru obţinerea producţiei din fiecare regiune, în ipoteza că s-ar lucra cu cheltuielile medii $\bar{t}_{\text{I}}$ şi $\bar{t}_{\text{II}}$.

Astfel, pentru regiunea *A* s-a obţinut un volum al producţiei de 18 798 om-ani (169,2.100 + 31,3.60), iar pentru regiunea *B*, un volum de 17 596 om-ani (169,2 . 30 + 31,3 . 400).

În sfîrşit, cu aceste date autorul calculează indicele teritorial al volumului fizic al producţiei (producţia regiunii *B* s-a luat ca bază):

$$I^{\text{ter.}}_{\text{vol. fiz.}}\left(\frac{A}{B}\right) = \frac{18\,798}{17\,596} = 1{,}068. \qquad \text{(XVI, 3.1)}$$

---

aplicarea la studiul mortalităţii a două grupe ale populaţiei şi anume: preoţii şi zeţarii. Din exemplul citat acolo (p. 124, pe baza datelor asupra populaţiei din Anglia, 1910) se pot observa unele rezultate paradoxale. Comparînd mortalitatea preoţilor cu cea a zeţarilor, rezultă că prima ar fi de aproape o dată şi jumătate mai mare decît a doua, deşi pe fiecare din cele două grupe de vîrstă, mortalitatea zeţarilor este considerabil mai mare. Acest rezultat paradoxal a putut să apară datorită deosebirilor mari în structura pe vîrste a celor două grupe ale populaţiei. Într-adevăr, din exemplul dat, rezultă că din grupa preoţilor majoritatea sînt bătrîni (peste 60% au vîrsta de peste 45 ani) pe cînd în grupa zeţarilor sînt mulţi tineri cu o mortalitate, în general, mai mică. Pentru înlăturarea influenţei variaţiei structurale pe grupe de vîrstă, se poate lua o structură „standard" (în exemplul folosit drept structură standard a fost luată structura pe vîrste a întregii populaţii de bărbaţi din Anglia, în 1931), în raport cu care se recalculează coeficientul mediu standard al mortalităţii celor două grupe de populaţie. De data aceasta rezultă evident că mortalitatea zeţarilor este cu mult mai ridicată decît a preoţilor.

Precizăm că metoda coeficienţilor standard se poate aplica în general la indicatori calitativi ca: natalitatea, preţul de cost, productivitatea muncii etc. după cum vom vedea mai departe.

Scanned with OKEN Scanner

După părerea lui S. Iughenburg, metoda coeficienților standard oferă o bază obiectivă pentru alegerea unui sistem de ponderi ai acestor indici și aplicarea ei va fi rațională în toate cazurile în care sistemul de ponderare folosit în mod obișnuit, pe baza structurii unuia din elementele comparate, nu poate fi considerat ca fiind fundamentat din punct de vedere economic.

Metoda lui L. M. Satunovski. Criticînd, parțial, concepția lui S. Iughenburg, L. M. Satunovski *) formulează cîteva principii care — după părerea sa — trebuie să stea la baza metodologiei de calcul al indicilor teritoriali.

L. M. Satunovski remarcă, în primul rînd, că problema alegerii bazei de comparație este inseparabil legată de problema sistemului de ponderi ai indicilor teritoriali ; că rezolvarea acestei din urmă probleme este în funcție nemijlocită de rezolvarea primei. Ignorarea acestei condiții primordiale a unei juste construiri a unui sistem de indici teritoriali constituie sursa principală de erori în problema alegerii ponderilor.

În al doilea rînd, după concepția acestui autor, problema comparării pe plan teritorial a diferiților indicatori se rezumă, în ultimă instanță, la necesitatea de a determina situația fiecărui obiect comparat, în raport cu *toate obiectele* din colectivitatea dată și nu în raport numai cu unul din ele ; că această necesitate apare în majoritatea cazurilor și că, deci, ea nu poate fi ignorată.

Prin urmare, tuturor obiectelor comparate trebuie să li se asigure condiții egale de comparabilitate. Or, asigurarea lor nu se poate face decît luînd drept bază de comparație întreaga colectivitate de obiecte comparate.

Indicii construiți pe această bază vor caracteriza rolul fiecărui obiect în formarea nivelului mediu al caracteristicii studiate sau a volumului respectiv, caracterizînd întreaga colectivitate de obiecte. Tocmai datorită acestui fapt, indicii teritoriali ai fiecărui obiect în parte din colectivitatea studiată devin comparabili și între ei.

Alegerea întregii colectivități ca bază de comparație — subliniază L. M. Satunovski — creează, în același timp, un criteriu obiectiv pentru construirea sistemului corespunzător de ponderi, respectîndu-se cu strictețe — și în domeniul indicilor teritoriali — principiile generale consacrate în teoria și practica statisticii sovietice :

— la construirea indicilor teritoriali privind indicatorii calitativi, operația de ponderare se va face cu ajutorul indicatorilor cantitativi (de structură) ai acelui obiect pentru care se calculează indicele respectiv și, invers,

— la construirea indicilor teritoriali privind indicatorii cantitativi, ponderarea se va face cu ajutorul nivelurilor medii ale indicatorilor calitativi, caracterizînd întreaga colectivitate de obiecte.

Ne vom opri acum asupra schemei teoretice propuse de L. M. Satunovski.

*) L. M. Satunovski, *K voprosu ob iscislenii teritorialnîh indeksov.* Vestnik Statistiki, 7 (1958).

Notația folosită :

$t_1, t_2, t_3, \ldots, t_n$, timpul efectiv cheltuit pentru fabricarea unui anumit produs în diferite întreprinderi (regiuni) ;

$q_1, q_2, q_3, \ldots, q_n$, cantitățile produsului respectiv, fabricate de diferite întreprinderi (regiuni) ;

$Q = q_1 + q_2 + \ldots + q_n$, cantitatea totală a produselor fabricate de întreaga colectivitate de întreprinderi (regiuni) ;

$\bar{t}$, timpul mediu cheltuit pe întreaga colectivitate de întreprinderi (regiuni) pentru fabricarea produsului respectiv.

În cadrul exemplului luat, autorul construiește următorul sistem de indici teritoriali, pentru fiecare întreprindere în parte : indicele volumului fizic al activității de producție ($I^{\text{ter.}}_{\text{vol. fiz.}}$) este egal cu produsul dintre indicele productivității muncii, incluzînd și influența deosebirilor de structură ($I^{\text{ter.}}_{\text{pr. m.}}$), și indicele timpului efectiv cheltuit de fiecare întreprindere ($I^{\text{ter.}}_{\text{chelt. timp}}$).

Sistemul poate fi concretizat sub forma indicilor agregați (ținînd seama că baza de comparație o reprezintă indicatorii întregii colectivități de întreprinderi și că aplicăm principiile de ponderare consacrate în statistica sovietică), de exemplu, pentru o întreprindere oarecare (I), în comparație cu totalitatea întreprinderilor :

$$I^{\text{ter.}}_{\text{vol. fiz.(I)}} = \frac{\Sigma \bar{t} q_1}{\Sigma \bar{t} Q}, \qquad \text{(XVI, 3.2)}$$

$$I^{\text{ter.}}_{\text{pr. m.(I)}} = \frac{\Sigma \bar{t} q_1}{\Sigma t_1 q_1}, \qquad \text{(XVI, 3.3)}$$

$$I^{\text{ter.}}_{\text{chelt. timp. (I)}} = \frac{\Sigma t_1 q_1}{\Sigma \bar{t} Q}, \qquad \text{(XVI, 3.4)}$$

Sistemul se verifică :

$$\frac{\Sigma \bar{t} q_1}{\Sigma \bar{t} Q} = \frac{\Sigma \bar{t} q_1}{\Sigma \bar{t}_1 q_1} \cdot \frac{\Sigma t_1 q_1}{\Sigma \bar{t} Q}. \qquad \text{(XVI, 3.5)}$$

După cum se vede, indicele volumului fizic al activității de producție*) este un indice cu ajutorul căruia se compară spațial (static) indicatorii cantitativi și, de aceea, drept ponderi au fost luați indicatorii calitativi (nivelurile medii ale cheltuielilor de timp de muncă pe unitatea de produs), calculați pe întreaga colectivitate de obiecte (întreprinderi), luată ca bază de comparație.

*) Satunovski folosește denumirea de indice al *activității* de producție și nu al volumului fizic al *producției*, întrucît indicele obținut — în speță pentru prima întreprindere — fiind calculat în raport cu volumul producției pe întreaga colectivitate de întreprinderi, măsoară nemijlocit *greutatea specifică* a întreprinderii, în volumul total al producției tuturor întreprinderilor (explicația dată de autor ; sublinierile noastre).

Prin conținutul său economic, acest indice caracterizează un fenomen complex, determinat, în esență, de doi factori principali: nivelul mediu al productivității muncii și volumul de cheltuieli de timp de muncă de la întreprinderea respectivă, ambele comparate, prin construcția indicilor corespunzători (indici factoriali), cu productivitatea medie și, respectiv, volumul total al cheltuielilor de timp de muncă, pe ansamblul colectivității studiate.

Observăm că și la construirea indicilor factoriali s-au respectat principiile de ponderare consacrate în statistica socialistă.

Mai observăm, apoi, că indicele factorial al productivității muncii $\left(\frac{\Sigma \bar{t} q_1}{\Sigma t_1 q_1}\right)$, comparînd indicatorii calitativi (productivitatea medie din întreprinderea noastră cu productivitatea medie pe ansamblul colectivității), a fost construit pe baza unor ponderi de indicatori cantitativi (indicatori de structură) ai întreprinderii comparate (q_1). Este evident că acest indice, la rîndul lui, este un rezultat al influenței variației în spațiu a doi factori și anume: nivelul diferit al productivității muncii (de la întreprindere la ansamblul colectivității de întreprinderi) și structura diferită a ponderilor (idem).

L. M. Satunovski consideră că uneori, în practică, s-ar putea ivi necesitatea de a fi calculat un indice teritorial al productivității muncii, eliberat de influența variației în spațiu a structurii.

În acest scop — susține autorul — s-ar putea lua structura medie a bazei de comparație (Q), construindu-se următorul indice pentru prima întreprindere din exemplul considerat:

$$I^{\text{ter.}}_{\text{pr. m.}_{(I)}} = \frac{\Sigma \bar{t} Q}{\Sigma t_1 Q}. \qquad \text{(XVI, 3.6)}$$

L. M. Satunovski precizează însă că, în general, influența reală pe care o exercită diferiți factori asupra fiecărui obiect în parte poate fi corect măsurată numai în cazul cînd în construcția indicilor respectivi este folosită structura efectivă a ponderilor. Tocmai în această problemă se manifestă mai mult dezacordul acestui autor față de concepția lui S. Iughenburg, care, recomandînd metoda coeficienților standard pentru calculul indicilor teritoriali, practic, limitează acest calcul la indicii teritoriali medii, eliberați de influența variației structurale.

Metoda lui G. Baklanov *). După părerea acestui autor problema elaborării indicilor teritoriali nu a fost rezolvată nici de S. Iughenburg și nici de L. Satunovski.

În primul rînd, G. Baklanov se pronunță împotriva măsurării volumului producției (sau a volumului activității productive) și implicit, împotriva măsurării productivității muncii, cu ajutorul *normelor medii de muncă pe ramură*. După părerea sa, volumul mediu efectiv de muncă pe ramură nu este un indicator potrivit pentru a caracteriza volumul activității productive, întrucît în acest mod este ignorat flagrant, gradul de

*) G. Baklanov, *Ob iscislenii teritorialnih indeksov*. Vestnik Statistiki, *11*, 34—54 (1958).

înzestrare tehnică a muncii. Adoptîndu-se acest procedeu — arată autorul — se poate ajunge la situaţii ca unele activităţi de producţie şi chiar ramuri întregi ale industriei cu un nivel mai scăzut de mecanizare a procesului de producţie să fie supraapreciate în mod artificial în ceea ce priveşte volumul activităţii de producţie şi invers, subapreciate întreprinderi şi ramuri industriale cu un înalt nivel de înzestrare tehnică. Este inadmisibil, deci, să comparăm volumul producţiei diferitelor întreprinderi, fără să ţinem seama de gradul de înzestrare tehnică a muncii.

Argumentele lui G. Baklanov în această privinţă sînt îndreptate îndeosebi, împotriva părerii lui S. Iughenburg care, după cum am văzut, propune ca — spre deosebire de metodologia existentă — volumul producţiei şi productivitatea muncii să fie calculate pe baza cheltuielilor medii de timp de muncă pe 1 000 ruble producţie globală (calculate pe ramură sau pe ansamblul economiei naţionale) şi — pe aceeaşi bază — să se calculeze apoi şi indicii teritoriali.

G. Baklanov critică în al doilea rînd şi sistemul propus de L. M. Satunovski care foloseşte *cheltuielile medii de muncă vie pe unitate de produs* drept unităţi de măsură pentru determinarea volumului producţiei.

Aceste unităţi de măsură nu pot asigura nici o idee justă asupra volumului producţiei pentru acelaşi motiv — cunoscut — că unele ramuri întregi ale industriei, cu un nivel mai redus de mecanizare a procesului de producţie, ar apărea în mod artificial cu un „volum fizic al producţiei" mai mare decît al celorlalte ramuri; pe de altă parte, aplicîndu-se acest procedeu, se ignorează complet munca trecută cuprinsă în producţia oricărei întreprinderi şi în plus munca vie se ia în considerare fără a se ţine seama de diferenţele ei calitative.

Sistemul indicilor teritoriali propus de L. Satunovski nu este acceptabil nici din punct de vedere practic, deoarece necesită calcule foarte greoaie. G.Baklanov arată că pentru a calcula indicele teritorial al volumului fizic al productivităţii muncii pentru o regiune dată, trebuie să fie primite, în prealabil, date complete asupra cantităţilor de produse fabricate (iar pentru produse complexe — asupra fabricării lor pe stadiile producţiei) şi asupra cheltuielilor de muncă pe produse, de la toate regiunile considerate. Fără aceste date nu se va putea obţine volumul mediu de muncă necesar ($\bar{t}$) pentru calculul indicilor teritoriali din sistemul lui L. Satunovski.

Deosebit de aceste motive, G. Baklanov mai arată că indicele teritorial al productivităţii muncii propus de L. Satunovski poate să formeze, în unele cazuri, o idee cu totul greşită asupra raportului între productivitatea muncii din regiunile comparate.

Fundamentîndu-şi propriul sistem de indici teritoriali, G. Baklanov formulează principiul concordanţei depline ce trebuie să existe între indicii dinamicii şi indicii teritoriali. Nu există decît o problemă unică a elaborării atît a indicilor teritoriali cît şi a indicilor dinamicii — arată autorul — şi de aceea propune ca evaluarea comparativă a volumului producţiei industriale atît în timp cît şi în spaţiu (pe întreprinderi, regiuni economice etc.), să se facă prin *compararea directă a unuia din următorii doi indicatori: producţia globală calculată după metoda economiei naţionale sau a volumului activităţii de producţie*, dedus prin diferenţa între costul producţiei globale

după metoda de întreprindere şi costul bunurilor materiale şi al serviciilor prestate de terţi). Folosirea acestei metode va face să fie eliminate o serie de particularităţi ale producţiei globale calculată după metoda de întreprindere şi care o fac improprie pentru caracterizarea dinamicii volumului producţiei şi a productivităţii muncii ca şi pentru construirea indicilor teritoriali.

De la indicatorii volumului de producţie se poate trece apoi uşor la calculul indicilor productivităţii muncii — indici ai dinamicii şi teritoriali.

În exemplul considerat de autor, în care se compară volumul producţiei globale (după metoda economiei naţionale) şi productivitatea pe un muncitor în două regiuni economice, G. Baklanov ajunge la concluzia că indicatorii productivităţii medii sînt influenţaţi considerabil de structura diferită a ramurilor din regiunile comparate.

Pentru a elimina influenţa deosebirilor în structura ramurilor din cele două regiuni şi G. Baklanov recurge la metoda coeficienţilor standard propusă de S. Iughenburg.

Cu totul altfel poate fi pusă problema — arată G. Baklanov — la stabilirea indicilor teritoriali ai *preţului de cost* al producţiei industriale (o asemenea problemă se pune, evident, numai cînd sînt mai multe tipuri de produse comparabile).

G. Baklanov consideră că pentru comparaţii spaţiale a nivelului preţului de cost ar fi indicat să se folosească cunoscuta formulă a indicelui preţului de cost după metoda pe ramură:

$$\frac{\Sigma q_1 z_1}{\Sigma q_1 \bar{z}_0}. \qquad \text{(XVI, 3.7)}$$

în care $\bar{z}_0$ reprezintă nivelul mediu pe ramură al preţului de cost pe unitate de produs în perioada de bază.

În fiecare regiune trebuie calculat indicele dinamicii preţului de cost după metoda pe ramură pentru producţia comparabilă. Numărătorul $\Sigma q_1 z_1$ va reprezenta, în acest caz, cheltuielile efective pentru producţia comparabilă din perioada curentă, a regiunii respective, iar numitorul $\Sigma q_1 \bar{z}_0$, cheltuielile convenţionale pentru aceeaşi producţie, calculată în preţul de cost mediu pe ramură din perioada de bază (pe ansamblul ţării). Caracterizarea comparativă a nivelului preţului de cost pe teritorii se va reduce apoi la compararea indicilor calculaţi după metoda arătată.

Evident, aplicarea acestei metode nu elimină influenţa structurii diferite pe regiuni a producţiei din perioada curentă. Pentru a elimina şi această influenţă, G. Baklanov propune tot metoda coeficienţilor standard, în care scop se pot considera cantităţile de produse pe scara întregii economii, după formula:

$$\frac{\Sigma q_{(\text{tot})_1} \cdot z_1}{\Sigma q_{(\text{tot})_1} \cdot \bar{z}_0} \qquad \text{(XVI, 3.8)}$$

în care $q_{(\text{tot})_1}$ reprezintă cantităţi de produse fabricate în perioada curentă în întreaga industrie.

Metoda lui L. Kazineț. L. Kazineț *) se alătură părerii formulate de S. M. Iughenburg și G. I. Baklanov în ceea ce privește înțelesul însuși al noțiunii de „indice teritorial" folosindu-l ca o mărime relativă care reflectă rezultatul comparării nivelului diferiților indicatori economici în spațiu, în special, din două regiuni economice.

L. Satunovski — arată L. Kazineț — substituie problema alcătuirii indicilor teritoriali cu problema comparării indicatorilor care se referă la o unitate a colectivității studiate cu indicatorii care se referă la întreaga colectivitate, rezolvînd deci o problemă de domeniul mărimilor relative și nu al indicilor teritoriali.

În esență, concepția lui L. Kazineț de elaborare a indicilor teritoriali — ca și a lui S. Iughenburg — se reazimă pe metoda coeficienților standard folosiți în demografie în cele două variante ale ei : *metoda directă* și *metoda indirectă.*

Prin analogie cu indicatorii demografici, L. Kazineț propune ca indicii teritoriali să fie calculați diferențiat, după natura indicatorilor considerați :

1°. *Indici teritoriali ai indicatorilor calitativi.* Autorul consideră în primul rînd indicele prețului de cost. Notînd cu Q numărul de unități ale producției de tip dat, care fac parte din colectivitatea standard și cu Q_1, Q_0 indicatorii producției, cu z_1 și z_0 indicatorii prețului de cost mediu al acestor produse, la două grupe de întreprinderi comparabile, iar z, prețul de cost standard, indicii teritoriali pot fi calculați astfel :

a) metoda directă,

$$I_{\text{ic.}}^{\text{ter.}} = \frac{\Sigma z_1 Q}{\Sigma z_0 Q}; \qquad \text{(XVI, 3.9)}$$

b) metoda indirectă,

$$I_{\text{ic.}}^{\text{ter.}} = \frac{\Sigma z_1 Q_1}{\Sigma z Q_1} : \frac{\Sigma z_0 Q_0}{\Sigma z Q_0}. \qquad \text{(XVI, 3.10)}$$

În al doilea rînd, autorul se referă și la indicii teritoriali ai indicatorilor calitativi ce se alcătuiesc pe baza valorilor inverse ale mărimii care urmează să fie exprimată prin indici. Astfel, indicii teritoriali ai productivității muncii la fabricarea unui anumit produs, alcătuiți după metoda directă și indirectă, vor avea următoarea formă :

$$I_{\text{pr. m.}}^{\text{ter.}} = \frac{\Sigma t_0 Q}{\Sigma t_1 Q} \quad \text{și} \quad I_{\text{pr. m.}}^{\text{ter}} = \frac{\Sigma t_0 Q_0}{\Sigma t Q_0} : \frac{\Sigma t_1 Q_1}{\Sigma t Q_1}, \qquad \text{(XVI, 3.11)}$$

în care t_1, t_0 reprezintă valorile corespunzătoare ale productivității muncii, iar t productivitatea medie generală (ponderea standard).

2°. *Indici teritoriali ai indicatorilor cantitativi.* La calculul indicelui teritorial al volumului fizic al producției, de exemplu, L. Kazineț arată că este necesar să luăm drept ponderi standard, cheltuielile medii de timp muncă aferente volumului standard al producției de tipul respectiv (ele

*) L. Kazineț, *O metodah postroeniia teritorialnîh indeksov.* Vestnik Statistiki, *3*, 58—66 (1959).

sînt notate cu t_i — pentru fiecare fel de produs), obținînd prin cele două variante următoarele forme de indici teritoriali:

a) *metoda directă*

$$I^{\text{ter.}}_{\text{vol. fiz.}} = \frac{\Sigma tQ_1}{\Sigma tQ_0}; \qquad \text{(XVI, 3.12)}$$

b) *metoda indirectă*

$$I^{\text{ter.}}_{\text{vol. fiz.}} = \frac{\Sigma t_1 Q_1}{\Sigma t_1 Q} : \frac{\Sigma t_0 Q_0}{\Sigma t_0 Q}. \qquad \text{(XVI, 3.13)}$$

L. Kazineț observă că indicii teritoriali calculați după metoda directă și cea indirectă sînt reversibili. De exemplu, indicele productivității muncii de la primul grup spațial (1) față de al doilea (0) este egal cu mărimea inversă a productivității muncii de la al doilea grup spațial (0) față de primul (1).

Se mai observă că dacă în formula (XVI, 3.11) vom schimba baza de comparație, indicii corespunzători ai productivității muncii vor avea următoarea formă:

$$\frac{\Sigma t_1 Q}{\Sigma t_0 Q} \text{ și } \frac{\Sigma t_1 Q_1}{\Sigma t Q_1} : \frac{\Sigma t_0 Q_0}{\Sigma t Q_0}, \qquad \text{(XVI, 3.14)}$$

reprezentînd mărimi inverse față de cele prezentate la (XVI, 3.11). Aceasta înseamnă — remarcă L. Kazineț — că în nici un caz concluziile cu privire la nivelul comparativ al indicatorului analizat la grupele spațiale respective, nu depind de alegerea bazei indicelui teritorial, indiferent de varianta de calcul folosită.

Pentru a calcula indicele teritorial al cheltuielilor de timp de muncă — indice derivat — autorul observă că el se poate calcula în două moduri și anume:

Se împarte indicele volumului fizic calculat după metoda directă la indicele productivității muncii calculat după metoda indirectă și se obține

$$I^{\text{ter.}}_{\text{chelt. timp.}} = \frac{\Sigma tQ_1}{\Sigma tQ_0} : \left(\frac{\Sigma t_0 Q_0}{\Sigma tQ_0} : \frac{\Sigma t_1 Q_1}{\Sigma tQ_1}\right) = \frac{\Sigma t_1 Q_1}{\Sigma t_0 Q_0}. \qquad \text{(XVI, 3.15)}$$

Același indice se obține prin împărțirea indicelui volumului fizic calculat prin metoda indirectă la indicele productivității muncii calculat prin metoda directă:

$$I^{\text{ter.}}_{\text{chelt. timp.}} = \left(\frac{\Sigma t_1 Q_1}{\Sigma t_1 Q} : \frac{\Sigma t_0 Q_0}{\Sigma t_0 Q}\right) : \frac{\Sigma t_0 Q}{\Sigma t_1 Q} = \frac{\Sigma t_1 Q_1}{\Sigma t_0 Q_0}. \qquad \text{(XVI, 3.16)}$$

După cum se vede, în sistemul indicilor teritoriali conceput de L. Kazineț se realizează o strînsă corelație: indicii teritoriali ai indicatorilor cantitativi, calitativi și derivați sînt legați reciproc prin formule adoptate în general, cu condiția necesară și suficientă ca indicii indicatorilor calitativi și cantitativi să fie calculați după metode diferite.

Scanned with OKEN Scanner

În ipoteza în care amîndoi acești indici sînt calculați după aceeași metodă — directă sau indirectă — corelația reciprocă dintre cei trei indici dispare.

Explicînd, mai departe, sensul acestei corelații, L. Kazineț atrage atenția că — după părerea sa — folosirea ponderilor constante la alcătuirea unuia din indici încalcă forma de legătură dintre cei trei indici și din această cauză ei încetează de a mai avea caracterul de indici parțiali, analitici, ca în cazul indicilor dinamicii, transformîndu-se în indicatori sintetici generali. Indicii teritoriali — în sistemul lui L. Kazineț — nu pot figura drept indicatori analitici și prin urmare nu sînt proprii pentru aprecierea influenței modificării diferiților factori asupra modificării fenomenului complex. În schimb, corelația dintre indicii teritoriali ai indicatorilor cantitativi, calitativi și derivați poate fi folosită cu rezultate destul de bune pentru calculul indirect al unuia din ei, cunoscîndu-se mărimea celorlalți doi. De pildă, indicele teritorial al productivității muncii se poate calcula pe baza indicilor teritoriali al volumului producției și al cheltuielilor de timp de muncă.

Indicii teritoriali din sistemul lui L. Kazineț, calculați după metoda directă au un sens economic bine determinat, ca al oricăror indici agregați. Calculați pentru colectivități de produse comparabile (standard), acești indici reprezintă *medii* ale indicilor individuali.

Cu totul altul este sensul indicilor teritoriali calculați după metoda indirectă. În esență, calculul lor este bazat pe ipoteza că raportul a doi indici de bază este egal cu indicele în lanț respectiv ceea ce este just numai în cazul cînd indicii în lanț sînt calculați pe baza ponderilor constante. Or, în cazul indicelui teritorial calculat prin metoda indirectă este vorba de indici în lanț, ce se calculează pe baza ponderilor variabile. În afară de aceasta, indicele calculat după metoda indirectă reprezintă o formă specifică a indicelui mediu, ceea ce poate să ducă la paradoxuri, arătînd că indicele general devine fie mai mare decît cel mai mare, fie mai mic decît cel mai mic din indicii individuali. Tocmai de aceea, metoda indirectă trebuie considerată — subliniază L. Kazineț — ca o metodă aproximativă de construire a indicilor teritoriali la care trebuie să se recurgă în cazul imposibilității calculului cu ajutorul metodei directe.

Metoda lui Z. Sidlik. Aspecte noi, interesante, în construirea indicilor teritoriali sînt relevate de statisticianul sovietic Z. Sidlik *).

În afară de indicii teritoriali ai volumului fizic al producției, ai productivității muncii și prețului de cost, acest autor se ocupă și de o altă problemă importantă a teoriei indicilor teritoriali și anume, de metodologia indicilor teritoriali ai *utilizării mijloacelor de producție fixe și a mijloacelor circulante normate.*

Din capul locului — criticînd unitățile de măsură pentru comensurarea producției eterogene, propuse de S. Iughenburg și L. Satunovski — acest autor se alătură părerii lui G. Baklanov ca fiind cea mai indicată din punct de vedere teoretic și practic. Indicatorul producției globale calculat după metoda economiei naționale oferă posibilitatea de a fi comparate

*) Z. Sidlik, Vestnik Statistiki, 8, 51—58 (1959).

direct, în expresie valorică, şi pe plan interregional, volumul producţiei, nivelul productivităţii muncii şi coeficienţii de utilizare ai mijloacelor fixe productive.

Z. Sidlik propune totuşi — în interesul unei mai mari exactităţi — ca din valoarea producţiei globale calculată după metoda economiei naţionale *să fie eliminată amortizarea mijloacelor fixe*, cunoscut fiind faptul că greutatea specifică a amortizării mijloacelor fixe şi a producţiei globale din diferite ramuri ale industriei şi din regiuni va rămîne inegală chiar şi după inventarierea şi reevaluarea mijloacelor fixe *) (în viitor această inegalitate se va accentua în ramurile şi regiunile industriale cu un înalt nivel de mecanizare şi automatizare a producţiei).

Spre deosebire de G. Baklanov, pentru calculul indicelui teritorial al preţului de cost al producţiei comparabile, Z. Sidlik propune următoarele formule:

$$I_{\text{r.c.}}^{\text{ter.}} = \frac{\Sigma q(u)_1 z_1}{\Sigma q(u)_1 \bar{z}_1(u)} \quad \text{sau} \quad \frac{\Sigma q(r)_1 z_1}{\Sigma q(r)_1 \bar{z}(r)} \qquad \text{(XVI, 3.17)}$$

în care simbolul u înseamnă mărimi unionale, iar r, cele regionale.

Se observă că la numitor apare preţul de cost mediu din perioada curentă unional sau regional ($\bar{z}_1(u)$) sau ($\bar{z}_1(r)$) ceea ce, după părerea autorului, sporeşte semnificaţia economică a comparaţiei realizată cu ajutorul indicelui teritorial propus. Autorul mai propune o metodă indirectă de calcul al indicilor teritoriali ai preţului de cost al întregii producţii marfă ca raport al indicilor teritoriali ai cheltuielilor de producţie la o rublă producţie marfă, considerînd — în contradicţie cu G. Baklanov — că acest indice este totuşi un indice al preţului de cost şi nu al rentabilităţii.

De notat că la calculul indicilor teritoriali ai volumului fizic al producţiei, productivităţii muncii şi ai utilizării fondurilor fixe (considerînd producţia globală — după metoda economiei naţionale — ce revine la 100 sau 1000 ruble fonduri fixe), autorul recomandă *compararea directă între regiuni a indicatorilor respectivi*.

În schimb, la calculul utilizării mijloacelor circulante, el recomandă metoda coeficienţilor standard, pornindu-se fie de la structura medie pe ramuri a industriei regiunilor comparate — ceea ce va fi mai corect — fie de la structura pe ramuri a întregei industrii.

Metoda lui P. Joffe. De problema unui sistem închegat şi unitar al indicilor teritoriali ai productivităţii muncii se ocupă în special autorul sovietic P. Joffe **).

În esenţă, concepţia acestui autor se apropie de cea a lui L. M. Satunovski. Este raţional — spune P. Joffe — ca drept bază de calcul pentru indicele teritorial al productivităţii muncii privind un produs oarecare, să se ia date pentru ramura corespunzătoare a economiei naţionale sau ndustriei, adică date pentru toate întreprinderile (sau gospodăriile) care

*) La 1 ianuarie 1960 a avut loc în U.R.S.S. inventarierea şi reevaluarea mijloacelor fixe din economia naţională.

**) P. Joffe, *O postroenii sisteme teritorialnth indeksov proizvoditelnosti truda*. Vestnik Statistiki, 2, 70—72 (1959).

produc produsul respectiv. Alegerea bazei de comparaţie soluţionează totodată problema măsurării consumului specific de muncă necesar pentru calculul indicatorilor productivităţii muncii pe întreprinderi, raioane, regiuni etc. Este evident că la calculul acestor indicatori, volumul producţiei trebuie să fie măsurat cu ajutorul normelor medii specifice pe ramură ale timpului de muncă vie efectiv cheltuită în perioada respectivă.

Prin urmare, volumul producţiei se va exprima prin formula

$$P = \Sigma \bar{t} q \qquad \text{(XVI, 3.18)}$$

în care q este cantitatea produsă, în expresie naturală, din produsul dat iar $\bar{t}$ volumul mediu efectiv de muncă vie pe ramură, pe o unitate din acest produs.

Notînd cu T fondul timpului de muncă consumat efectiv pentru obţinerea produsului respectiv şi cu k indicatorul productivităţii muncii pentru fiecare produs, vom avea

$$k = \frac{\bar{t} q}{T}. \qquad \text{(XVI, 3.19)}$$

Mărimea acestui indicator arată cîte ore medii de muncă pe ramură revin pe o unitate a timpului de lucru la fabricarea produsului dat. Dacă ţinem seama de faptul că $\bar{t} = \frac{1}{\bar{w}}$ ($\bar{w}$ fiind producţia medie pe ramură obţinută într-o unitate de timp de muncă) şi că deci $q = w \cdot T$, indicatorul productivităţii muncii k poate fi exprimat şi în felul următor:

$$k = \frac{\bar{t} q}{T} = \frac{\frac{w\,T}{\bar{w}}}{T} = \frac{w}{\bar{w}}. \qquad \text{(XVI, 3.20)}$$

După cum se vede, acest indicator exprimă nivelul relativ al productivităţii muncii.

Ponderînd indicii productivităţii muncii calculaţi pe întreprinderi cu fondurile corespunzătoare ale timpului de lucru sau cu numărul muncitorilor, se obţin indicii teritoriali sintetici pe raioane, regiuni etc. de forma

$$I^{\text{ter.}}_{\text{sintetic}} = \frac{\Sigma k T}{\Sigma T}. \qquad \text{(XVI, 3.21)}$$

Această formulă de bază a indicelui teritorial al productivităţii muncii corespunde în esenţă cu formula din sistemul indicilor teritoriali propus de L. Satunovski

$$I^{\text{ter}}_{\text{p.m.}} = \frac{\Sigma \bar{t} q}{\Sigma t q}.$$

Într-adevăr, cunoscînd că : $\bar{t} = \frac{1}{\bar{w}}$; $q = T \cdot w$ și $k = \frac{w}{\bar{w}}$, formula lui L. Satunovski capătă forma matematică a celei propuse de P. Joffe :

$$I_{\text{p.m.}}^{\text{ter.}} = \frac{\Sigma \bar{t} q}{\Sigma t q} = \frac{\Sigma \frac{1}{\bar{w}} T \cdot w}{\Sigma T} = \frac{\Sigma \frac{w}{\bar{w}} T}{\Sigma T} = \frac{\Sigma k T}{\Sigma T}. \qquad \text{(XVI, 3.21a)}$$

Folosind formula de bază (XVI,3.21), P. Joffe construiește un sistem complex de indici teritoriali ai productivității muncii, diferențiat pe tipuri principale.

Diferite tipuri de indici oglindesc influența tuturor factorilor productivității muncii fără excepție, inclusiv influența schimbărilor structurale. Autorul deosebește, astfel, următoarele feluri de indici teritoriali :

1) indici de produse ; 2) indici generali pe întreprinderi cu producție eterogenă ;

3) indici sintetici pe regiuni și republici ;

4) indici teritoriali de comparare pentru diferite întreprinderi și regiuni.

Indicii de tipul 1) și 2) sînt cei ce se obțin prin formulele (XVI, 3.20) și (XVI, 3.21). Prin formula (XVI, 3.21) se obține și indicele teritorial sintetic pe regiuni sau republică ; el este în acest caz un indice mediu aritmetic pentru totalitatea întreprinderilor și a produselor, calculat pe baza structurii efective, prin înregistrarea totală sau selectivă.

În grupa indicilor teritoriali de tipul 3), autorul mai propune un indice teritorial sintetic *eliberat de influența deosebirilor în structură* și care este identic cu indicele corespunzător propus de L. Satunovski de forma $\frac{\Sigma \bar{t} Q}{\Sigma t Q}$ (unde Q este cantitatea totală a producției produsului dat pentru toate întreprinderile incluse în cercetare).

Formula lui P. Joffe este următoarea :

$$I_{\text{infl,str.}}^{\text{ter.}} = \frac{\Sigma F_t}{\Sigma \frac{1}{k} F_t}, \qquad \text{(XVI, 3.22)}$$

în care F_t reprezintă fondul total al timpului de lucru cheltuit pentru a obține întreaga producție de același fel sau eterogenă.

Cît privește indicii teritoriali de tipul 4) — de comparare — pe întreprinderi și regiuni, P. Joffe propune cinci feluri de asemenea indici, care sînt combinații ale primelor trei *).

Metoda lui V. Peregudov. Deosebit de interesantă este concepția statisticianului sovietic V. Peregudov **).

*) Pentru documentare a se vedea articolul citat la pag. 436, p. 80, tabela 3, în care sînt date formulele respective și exemplele de calcul.

**) V. Peregudov, *K voprosu o sravnenii po raionam kacestvennih pokazatelei.* Vestnik Statistiki, 5, 45—55 (1959).

Examinînd particularitatea specifică a comparării între regiuni a indicatorilor economici, V. Peregudov constată că — spre deosebire de comparările în timp — aici sînt legitime în egală măsură comparările în ambele sensuri; regiunile comparate stau, cum s-ar zice, alături şi nici una din ele nu poate avea pretenţia de a fi considerată în mod categoric „de bază" sau — prin analogie cu comparările în timp — drept „perioadă curentă". În comparările spaţiale ale indicatorilor economici apare, deci, cerinţa „reversibilităţii' şi ea se referă nu numai la indicele luat în întregime, ci şi la elementul care se compară cu ajutorul indicelui respectiv. Această constatare l-a dus, evident, la contradicţia pe care a relevat-o S. Iugenburg şi alţi autori care scot în evidenţă faptul că regiunile comparate „sînt una mai bună decît alta" dacă indicii teritoriali respectivi sînt construiţi luîndu-se drept bază de comparaţie mai întîi o regiune şi apoi alta şi respectîndu-se schema de ponderare acceptată în general în statistica socialistă.

Rezultatele apar diferite şi contradictorii tocmai datorită faptului că în regiunile comparate structura producţiei este deosebită.

V. Peregudov subliniază că, aparent, soluţia problemei poate fi găsită şi contradicţia lichidată dacă drept pondere la construirea indicelui teritorial se ia volumul total al producţiei celor două regiuni comparate luate împreună (ştim că aceasta este de fapt soluţia pe care a dat-o L. M. Satunovski şi P. Joffe). Indicii teritoriali ai indicatorilor calitativi construiţi în acest mod ar elimina influenţa deosebirilor de structură, produsul lor fiind în mod obligatoriu egal cu unitatea.

O cercetare mai adîncă a problemei, arată V. Peregudov, dezvăluie, însă, noi contradicţii, vădind inconsistenţa acestei soluţii.

Exemplele pe care le examinează scot în evidenţă faptul că asupra indicelui teritorial al indicatorului calitativ considerat (preţul de cost) exercită influenţă trei factori şi anume: raportul de mărime dintre nivelurile preţului de cost din regiunile comparate; structura producţiei pe regiuni; *greutatea specifică a regiunilor în volumul total al producţiei.*

Este evident că în practică nu se poate merge atît de departe — spune Peregudov — încît la calculul indicilor să se caute corective prin care să fie eliminată şi influenţa ultimului factor. Este mai raţional ca indicele preţului de cost să fie calculat fără a ţine seama de greutatea specifică a regiunilor comparate în volumul total al producţiei. Or, metoda de calcul bazată pe folosirea volumului total al producţiei nu satisface această condiţie şi de aceea nu soluţionează problema propusă.

După cum am văzut, contradicţia de bază constă în aceea că în cazul cînd se iau indicii individuali pentru diferite produse, calculul lor pentru două regiuni comparate ne duce la numere inverse al căror produs evident este egal cu unitatea. În schimb, în multe cazuri cînd se calculează indicii agregaţi care se referă la o grupă de diferite produse, această proprietate de „reversibilitate" dispare.

Peregudov constată totuşi că lipsa proprietăţii de „reversibilitate" la indicii agregaţi nu este o regulă generală. Sînt numeroase cazuri de indici agregaţi care păstrează această proprietate.

Scanned with OKEN Scanner

De exemplu, nu este greu de observat că proprietatea de „reversibilitate" se va păstra în toate cazurile cînd indicii individuali ai volumului fizic

$$i_q = \frac{q_2}{q_1}$$

sînt constanți, adică $i_q = k$. De asemenea, ea se va păstra și în cazul cînd toți indicii individuali ai prețului de cost sînt constanți:

$$i_c = \frac{c_2}{c_1} = k'.$$

Caracterul constant al indicilor individuali i_q și i_c nu este însă o condiție necesară și suficientă care poate să asigure „reversibilitatea" indicilor agregați. În practică se ivesc numeroase cazuri cînd deși i_q și i_c nu rămîn constanți, „reversibilitatea" indicilor agregați este totuși asigurată.

Printre altele, V. Peregudov examinează și exemplul din tabelul 76.

Tabelul 76

Feluri de produse	Regiunea 1			Regiunea 2			Cheltuieli de producție (convențională)	
	c_1	q_1	c_1q_1	c_2	q_2	c_2q_2	c_1q_2	c_2q_1
A	5	440	2 200	2,5	572	1 430	2 860	1 100
B	10	160	1 600	8	153,6	1 228,8	1 536	1 280
C	20	61	1 220	20	61	1 220	1 220	1 220
D	40	40	1 600	50	43,6	2 180	1 744	2 000
E	50	44	2 220	100	52,8	5 280	2 640	4 400
Total	—	—	8 820	—	—	11 338,8	10 000	10 000

În acest caz, indicii teritoriali ai prețului de cost, luînd alternativ drept bază de comparație fiecare regiune și folosind sistemul de ponderare în general admis, vor fi egali cu:

$$I_{2/1} = \frac{\Sigma c_2 q_2}{\Sigma c_1 q_2} = \frac{11\,338{,}8}{10\,000} = 1{,}134 \quad \text{și} \quad I_{1/2} = \frac{\Sigma c_1 q_1}{\Sigma c_2 q_1} = \frac{8\,820}{10\,000} = 0{,}882$$

sau schimbînd ponderile:

$$I_{2/1} = \frac{\Sigma c_2 q_1}{\Sigma c_1 q_1} = \frac{10\,000}{8\,820} = 1{,}134 \quad \text{și} \quad I_{1/2} = \frac{\Sigma c_1 q_2}{\Sigma c_2 q_2} = \frac{10\,000}{11\,338{,}8} = 0{,}882.$$

Pe baza exemplelor considerate, Peregudov ajunge la concluzia că pentru asigurarea „reversibilității" indicilor agregați există o condiție generală și anume, *independența sistemului indicilor* i_c *față de sistemul indicilor* i_q.

Or, din punct de vedere teoretic — arată Peregudov — este admisibil un singur criteriu pentru asigurarea condiției de independență a două mărimi x și y, constînd în egalitatea :

$$\overline{x \cdot y} = \bar{x} \cdot \bar{y}.$$

În cazul indicilor medii i_c și i_q care sînt mărimi medii ponderate, este necesar ca atît aceste medii cît și produsul lor mediu să se calculeze folosind ponderi corespunzătoare. Aceste ponderi însă — egale pentru toate cele trei medii — trebuie alcătuite pornind de la condițiile concrete și de la conținutul problemei.

Folosind notația : $i_c = \frac{c_2}{c_1}$ și $i_q = \frac{q_2}{q_1}$, iar pentru mărimile inverse : $i'_c = \frac{c_1}{c_2}$; $i'_q = \frac{q_1}{q_2}$ și pornind de la raportul cheltuielilor efective de producție $w_1 = c_1 q_1$; $w_2 = c_2 q_2$, autorul demonstrează că este îndeplinită condiția independenței lui i_c și i_q (sau ceea ce este același lucru — a lui i'_c și i'_q) și prin urmare, este îndeplinită condiția „reversibilității" indicilor agregați ai prețului de cost, calculați pentru compararea regiunilor, cînd avem egalitatea :

$$I_{(c,q)} = \frac{\Sigma c_2 q_2}{\Sigma c_2 q_1} \cdot \frac{\Sigma c_1 q_1}{\Sigma c_1 q_2} = 1. \qquad \text{(XVI, 3.23)}$$

Raționamentul demonstrației este următorul :

$$\frac{\Sigma c_2 q_2}{\Sigma c_1 q_1} = \frac{\Sigma \left(\frac{c_2 q_2}{c_1 q_1}\right) c_1 q_1}{\Sigma c_1 q_1} = \frac{\Sigma i_c i_q w_1}{\Sigma w_1} = \overline{i_c i_q},$$

$$\frac{\Sigma c_1 q_1}{\Sigma c_2 q_2} = \frac{\Sigma \left(\frac{c_1 q_1}{c_2 q_2}\right) c_2 q_2}{\Sigma c_2 q_2} = \frac{\Sigma i'_c i'_q w_2}{\Sigma w_2} = \overline{i'_c i'_q}.$$

Dacă indicii i_c și i_q se iau drept ponderi, trebuie luate cheltuielile de producție din prima regiune, $c_1 q_1$, și invers, dacă se iau indicii i'_c și i'_q drept ponderi, trebuie luate cheltuielile de producție din cea de-a doua regiune, $c_2 q_2$.

Dublul caracter de soluționare a problemei ponderilor va dispare, ambele variante contopindu-se vor forma un singur criteriu de independență i_c și i_q.

Dacă pentru compararea produsului mediu $\overline{i_c i_q}$ cu produsul mediilor $\bar{i}_c$ și $\bar{i}_q$ vom lua ca bază de comparație prima regiune, este necesar ca aceasta din urmă să fie ponderată cu $w_1 = c_1 q_1$, adică să le obținem după formulele :

$$\frac{\Sigma i_c w_1}{\Sigma w_1} = \frac{\Sigma \left(\frac{c_2}{c_1}\right) c_1 q_1}{\Sigma c_1 q_1} = \frac{\Sigma c_2 q_1}{\Sigma c_1 q_1}$$

și

$$\frac{\Sigma i_q w_1}{\Sigma w_1} = \frac{\Sigma\left(\frac{q_2}{q_1}\right) c_1 q_1}{\Sigma c_1 q_1} = \frac{\Sigma c_1 q_2}{\Sigma c_1 q_1}.$$

Prin urmare, în această variantă de ponderare independența între i_c și i_q va exista atunci cînd vom avea egalitatea

$$\overline{i_c i_q} = \bar{i}_c . \bar{i}_q .$$

Această egalitate o vom scrie sub forma

$$\frac{\Sigma i_c i_q w_1}{\Sigma w_1} = \frac{\Sigma i_c w_1}{\Sigma w_1} \cdot \frac{\Sigma i_q w_1}{\Sigma w_1}$$

sau

$$\frac{\Sigma c_2 q_2}{\Sigma c_1 q_1} = \frac{\Sigma c_2 q_1}{\Sigma c_1 q_1} \cdot \frac{\Sigma c_1 q_2}{\Sigma c_1 q_1}$$

sau

$$\left(\frac{\Sigma c_2 q_2}{\Sigma c_1 q_1}\right) : \left(\frac{\Sigma c_2 q_1}{\Sigma c_1 q_1} \cdot \frac{\Sigma c_1 q_2}{\Sigma c_1 q_1}\right) = \frac{\Sigma c_2 q_2 \cdot \Sigma c_1 q_1}{\Sigma c_2 q_1 \cdot \Sigma c_1 q_2} = 1.$$

În cazul celei de-a doua variante, ponderîndu-se mărimile inverse i_c' și i_q', vom avea

$$\frac{\Sigma i_c' i_q' w_2}{\Sigma w_2} = \frac{\Sigma i_c' w_2}{\Sigma w_2} \cdot \frac{\Sigma i_q' w_2}{\Sigma w_2}$$

în care

$$\frac{\Sigma i_c' w_2}{\Sigma w_2} = \frac{\Sigma\left(\frac{c_1}{c_2}\right) c_2 q_2}{\Sigma c_2 q_2} = \frac{\Sigma c_1 q_2}{\Sigma c_2 q_2}$$

și

$$\frac{\Sigma i_q' w_2}{\Sigma w_2} = \frac{\Sigma\left(\frac{q_1}{q_2}\right) c_2 q_2}{\Sigma c_2 q_2} = \frac{\Sigma c_2 q_1}{\Sigma c_2 q_2}.$$

Prin urmare

$$\frac{\Sigma c_1 q_1}{\Sigma c_2 q_2} = \frac{\Sigma c_1 q_2}{\Sigma c_2 q_2} \cdot \frac{\Sigma c_2 q_1}{\Sigma c_2 q_2}$$

sau

$$\left(\frac{\Sigma c_1 q_1}{\Sigma c_2 q_2}\right) : \left(\frac{\Sigma c_1 q_2}{\Sigma c_2 q_2} \cdot \frac{\Sigma c_2 q_1}{\Sigma c_2 q_2}\right) = \frac{\Sigma c_2 q_2 \cdot \Sigma c_1 q_1}{\Sigma c_2 q_1 \cdot \Sigma c_1 q_2} = 1.$$

Indicele din formula (XVI, 3.23) Peregudov îl denumește „indicele legăturii reciproce". La compararea indicatorilor economici calitativi pe

regiuni, „reversibilitatea" indicilor agregați se realizează dacă și numai dacă indicele legăturii reciproce $I_{(c,q)}$ este egal cu 1.

Această condiție o satisfac și datele din tabelul 76 în care i_c și i_q nu sînt constanți și totuși nu există dependență între ei:

$$I_{(c,q)} = \frac{11\,338{,}8 \cdot 8\,820}{10\,000 \cdot 10\,000} = 1.$$

Din punct de vedere al conținutului său economic, indicele legăturii reciproce $I_{(c,q)}$ reprezintă *specializarea relativă* a regiunilor. Este vorba de o specializare în sensul îngust al cuvîntului, în raport cu o anumită caracteristică. De pildă, nu putem vorbi de specializarea relativă a două regiuni alăturate (amîndouă producătoare de bumbac), dacă structura producției lor este aceeași, adică dacă $I_{(c,q)} = 1$.

În practică, însă, se întîlnesc frecvent regiuni cu specializare relativă (cînd $I_{(c,q)} \neq 1$) și problema care se pune este aceea de a găsi posibilitatea comparării și în aceste cazuri.

În acest scop Peregudov introduce media geometrică a indicelui legăturii reciproce

$$I_{\text{geom.}} = \sqrt{\frac{\Sigma c_2 q_2}{\Sigma c_1 q_2} \cdot \frac{\Sigma c_1 q_1}{\Sigma c_2 q_1}}, \qquad \text{(XVI, 3.24)}$$

cu care trebuie comparați indicii agregați obținuți pentru regiunile comparate.

Modul cum trebuie construiți în mod just indicii teritoriali, Peregudov îl ilustrează prin exemplul din tabelul 77.

Tabelul 77

Felurile de produse	Regiunea 1			Regiunea 2			Cheltuieli de producție (convențională)	
	c_1	q_1	c_1q_1	c_2	q_2	c_2q_2	c_1q_2	c_2q_1
A	5	440	2 200	2,5	600	1 500	3 000	3 100
B	10	160	1 600	8	175	1 400	1 750	1 280
C	20	61	1 220	20	60	1 200	1 200	1 220
D	40	40	1 600	50	36	1 800	1 440	2 000
E	50	44	2 200	100	25	2 500	1 250	4 400
Total	—	—	8 820	—	—	8 400	8 640	10 000

Indicii agregați respectivi, luînd drept bază de comparație mai întîi prima regiune (cu ponderile regiunii a doua) și apoi a doua (cu ponderile primei regiuni), vor fi:

$$I_{2/1} = \frac{\Sigma c_2 q_2}{\Sigma c_1 q_2} = \frac{8\,400}{8\,640} = 0{,}972,$$

$$I_{1/2} = \frac{\Sigma c_1 q_1}{\Sigma c_2 q_1} = \frac{8\,820}{10\,000} = 0{,}882.$$

Dacă vom compara acești doi indici cu unitatea, ajungem la concluzia absurdă că a doua regiune este mai bună decît prima (aici prețul de cost este cu 2,8% mai redus decît la prima) și în același timp, prima regiune este mai bună decît a doua (aici prețul de cost este mai scăzut cu 11,8%). Dacă însă vom calcula nivelul specializării relative a regiunilor folosind indicele geometric din formula (XVI, 3.24), vom obține

$$I_{\text{geom.}} = \sqrt{0{,}972 \cdot 0{,}882} = 0{,}926.$$

Rezultatul comparării indicilor $I_{2/1}$ și $I_{1/2}$ cu indicele specializării relative ne arată că a doua regiune are rezultate mai puțin favorabile în ceea ce privește prețul de cost, comparativ cu prima regiune.

Într-adevăr,

$$I_{2/1} = 0{,}972 > 0{,}926 \text{ iar } I_{1/2} = 0{,}882 < 0{,}926.$$

Influența specializării relative poate fi eliminată prin împărțirea indicilor $I_{2/1}$ și $I_{1/2}$ la indicele specializării relative.

Se obține astfel alți doi indici $I'_{2/1}$ și $I'_{1/2}$:

$$I'_{2/1} = I_{2/1} : I_{\text{geom.}} = 0{,}972 : 0{,}926 = 1{,}050,$$
$$I'_{1/2} = I_{1/2} : I_{\text{geom.}} = 0{,}882 : 0{,}926 = 0{,}952.$$

Indicii $I'_{2/1}$ și $I'_{1/2}$ sînt echivalenți cu indicii în care specializarea lipsește și de aceea produsul lor este egal cu 1:

$$I'_{2/1} \cdot I'_{1/2} = 1{,}05 \cdot 0{,}952 = 1.$$

În forma generală, indicii $I'_{2/1}$ și $I'_{1/2}$ capătă următoarea expresie:

$$I'_{2/1} = I_{2/1} : I_{\text{geom.}} = \frac{\Sigma c_2 q_2}{\Sigma c_1 q_2} : \sqrt{\frac{\Sigma c_2 q_2 \cdot \Sigma c_1 q_1}{\Sigma c_1 q_2 \cdot \Sigma c_2 q_1}} = \sqrt{\frac{\Sigma c_2 q_2 \cdot \Sigma c_2 q_1}{\Sigma c_1 q_2 \cdot \Sigma c_1 q_1}}$$

și

$$I'_{1/2} = I_{1/2} : I_{\text{geom.}} = \frac{\Sigma c_1 q_1}{\Sigma c_2 q_1} : \sqrt{\frac{\Sigma c_2 q_2 \cdot \Sigma c_1 q_1}{\Sigma c_1 q_2 \cdot \Sigma c_2 q_1}} = \sqrt{\frac{\Sigma c_1 q_1 \cdot \Sigma c_1 q_2}{\Sigma c_2 q_1 \cdot \Sigma c_2 q_2}}. \qquad \text{(XVI, 3.25)}$$

După cum se vede — și Peregudov remarcă acest lucru — formulele (XV, 3.25) sînt de tipul indicelui „ideal" al lui Irving Fisher care a considerat și el necesitatea de a fi satisfăcută cerința „reversibilității" într-un caz mai general și mai ales la indicii dinamicii, denumind-o „test al reversibilității în timp".

V. Peregudov atrage atenția că în cazul analizei fenomenelor social-economice în timp, această cerință nu are nici un sens real, iar teoria statisticii sovietice consideră în mod just că ea este pur formală. Nu aceeași situație este însă, în cazul comparării în spațiu a indicatorilor economici calitativi. Cît timp regiunile comparate sînt echivalente, „reversibilitatea în spațiu" este necesară deoarece în caz contrar nu vom putea ieși niciodată din cunoscuta contradicție, semnalată și de alți autori.

4. Observații și concluzii

Am scos în evidență, credem, problemele esențiale ce se ridică la elaborarea indicilor teritoriali în lumina concepțiilor noi, interesante, pe care le-am expus. Soluțiile metodologice propuse vădesc un considerabil progres în dezvoltarea multilaterală a teoriei indicilor teritoriali, conturarea unor principii directoare, științific fundamentate, pentru multiplele aplicații practice.

Desigur, numai verificarea practică a soluțiilor optime preconizate de autorii citați în principalele probleme ale teoriei indicilor teritoriali va constitui criteriul suprem al adevărului pe care-l conțin și al exactității științifice a concluziilor ce s-au desprins la sfîrșitul discuțiilor purtate. Majoritatea autorilor au și propus chiar metodele pe care le-au elaborat să fie experimentate practic și apoi extinse eventual în cadrul unor aplicații generale.

După cum am putut remarca, prin conținutul lor de idei, părerile formulate de autorii citați au unele elemente comune ceea ce ne permite să le grupăm în jurul a trei concepții de bază diferite.

În primul rînd, o deplină identitate de concepție și de metodologie statistică se remarcă la L. M. Satunovski și P. Joffe, deosebirile — neesențiale — referindu-se doar la forma matematică de construire a indicilor teritoriali și la legătura de sistem a acestora. De concepția acestor autori, și, în deosebi, de varianta lui L. M. Satunovski, se apropie — în unele privințe — și concepția lui L. Kazineț.

În al doilea rînd, elemente comune în ceea ce privește metodologia de calcul al diferiților indici teritoriali, găsim la S. Iughenburg, G. Baklanov, L. Kazineț și Z. Sidlik, deși în multe probleme importante, îndeosebi în cele referitoare la conținutul economic al indicatorilor propuși, părerile lor diferă substanțial.

În al treilea rînd, o părere deosebită o are V. Peregudov a cărui concepție am expus-o pe larg, considerînd-o ca fiind deosebit de interesantă cu toate că acest autor abordează numai problemele speciale legate de construirea indicilor teritoriali privind indicatorii calitativi.

Sintetizînd problemele puse în discuție, vom încerca să facem pe marginea lor cîteva considerații proprii.

1°. Vom face în primul rînd două constatări prealabile :

Majoritatea autorilor au abordat și au rezolvat în chip diferit, problema comensurabilității fenomenelor social-economice supuse comparării în spațiu deși — după părerea noastră — ea nu este o problemă specifică teoriei indicilor teritoriali, ci o problemă generală a statisticii economice. Cu aceeași stringență se pune această problemă, de exemplu, în studiul dinamicii indicatorilor economici cum sînt volumul producției globale, productivitatea muncii, prețul de cost ca și al altor fenomene și procese social-economice complexe ce nu se pot însuma direct.

Se constată apoi — și aceasta la toți autorii — preocuparea specială de a face o delimitare științifică riguroasă între indicii teritoriali ai indicatorilor *cantitativi* și indicii teritoriali ai indicatorilor *calitativi*, unii autori (P. Joffe, V. Peregudov) acordînd atenție deosebită acestora din urmă.

Această preocupare s-a impus, evident, datorită conținutului economic *specific* pe care-l au aceste două categorii statistice, datorită *sensului* lor logic și a *funcțiilor* analitice deosebite pe care trebuie să le îndeplinească.

Dacă ne referim la problema comensurabilității și la soluțiile propuse pentru rezolvarea ei, în cazul special al indicilor teritoriali, atrage atenția în mod deosebit părerea exprimată de G. Baklanov, și susținută de Z. Sidlik. În etapa actuală de dezvoltare a evidenței și statisticii, metoda propusă de G. Baklanov este într-adevăr cea mai potrivită.

Considerarea în calculul indicilor teritoriali a indicatorului producției globale calculat după metoda economiei naționale (sau a indicatorului privind volumul activității productive), are darul să înlăture particularitățile folosirii în acest scop a indicatorului producției globale calculate după metoda de întreprindere care implică — după cum se știe — o serie de neajunsuri în studiul seriilor în timp și în spațiu, asigurînd compararea *directă*, în expresie valorică, a volumului producției, a productivității muncii și altor indicatori economici.

Prin aplicarea acestei metode se soluționează în bună parte — după cum pe bună dreptate remarcă Z. Sidlik — problema metodelor de exprimare a deosebirilor existente în structura pe ramuri a producției din diferitele regiuni economice, în nivelul mecanizării și automatizării procesului de producție, în legăturile de cooperare productivă. De asemenea, se soluționează implicit, spinoasa problemă a alegerii bazei de comparație și se evită totodată, folosirea metodei coeficienților standard în cele două variante ale ei, ceea ce simplifică considerabil procesul comparațiilor pe plan interregional, problema construirii, în acest scop, a indicilor teritoriali.

Un alt argument puternic care pledează în favoarea metodei propuse de G. Baklanov, este de ordin practic. Din punct de vedere al activității practice, metoda lui G. Baklanov este mai acceptabilă de cît oricare alta, întrucît în fiecare întreprindere și în fiecare direcție de statistică se găsesc toate datele necesare pentru calculul indicatorilor producției globale după metoda economiei naționale sau a volumului activității de producție. Este totuși necesar să menționăm că un asemenea calcul ar necesita un efort suplimentar, știut fiind că în prezent producția globală se calculează după metoda de întreprindere. În afară de aceasta, traducerea în viață a metodei propusă de G. Baklanov ar întîmpina și o serie de dificultăți legate de funcționarea sistemului prețurilor, de structură a acestora, de nivelul rentabilității diferitelor produse etc. pe care nu este locul să le examinăm aici.

Credem că în cazul cînd la baza calculului indicilor teritoriali privind producția, productivitatea muncii, utilizarea fondurilor fixe etc. s-ar adapta indicatorul exprimînd *volumul producției globale după metoda economiei naționale*, ar trebui acceptată și propunerea lui Z. Sidlik ca acestui indicator să i se aducă o corecție referitoare la valoarea de amortizare a fondurilor fixe, o asemenea metodologie fiind în măsură să asigure calculelor un grad de precizie mai mare. Această soluție se impune cu atît mai mult cu cît este îndoielnic dacă scopul urmărit prin adăugarea mărimii amortizărilor la producția netă — acela ca la comparări în timp și spațiu

Scanned with OKEN Scanner

să se țină seama și de gradul diferit de înzestrare tehnică a muncii în întreprinderile industriale — va fi practic atins.

Într-adevăr, se știe că înzestrarea tehnică a muncii în forma ei mai generală — așa cum observă și L. Kazineț — este determinată de raportul dintre valoarea fondurilor fixe și numărul muncitorilor. Or, în formula lui G. Baklanov, fondurile fixe sînt rupte de numărul muncitorilor și nici nu se iau în întregime, ci în fracții corespunzătoare cotelor de amortizare care nu sînt aceleași pe ramuri industriale.

Avantajul considerabil al metodei propuse de G. Baklanov — în corecția lui Z. Sidlik — constă în faptul esențial că problema comensurabilității fenomenelor social-economice complexe se rezolvă în funcție de sistemul actual de evidență și statistică în care predomină indicatorii în expresie valorică și în consecință eficacitatea ei poate fi practic verificată.

În dezvoltarea economică de perspectivă a țărilor socialiste pe calea trecerii treptate la comunism, cînd progresul tehnic general asigură și statisticii cele mai moderne mijloace de calcul, se va aplica și extinde pe scară tot mai largă metoda *directă* de măsurare a valorii, în unități de timp de muncă. Această metodă este cea mai corectă, prezintă toate avantajele pe care le oferă expresia bănească a valorii și elimină, totodată, deficiențele acesteia.

Examinînd metodologia propusă de S. Iughenburg și L. M. Satunovski susținută apoi și de P. Joffe, am avut prilejul să constatăm că acești autori calculează volumul producției eterogene și toți indicatorii legați de acesta, în unități de timp de muncă cheltuită efectiv (om-ani), construind pe această bază și indicii teritoriali. Desigur că această metodă nu poate fi identificată cu metoda generală a viitorului în care întreaga activitate productivă să fie evaluată în norme medii de timp de muncă, ea totuși o anticipează și, fiind de aceeași esență, este fundamentată din punct de vedere științific. Însuși K. Marx, calculînd dinamica productivității muncii din Anglia în anii 1770—1840, a evaluat volumul maxim posibil al întregii producții din Anglia în om-ani muncă vie. O asemenea metodă de măsurare a volumului producției eterogene ține seama pe deplin, de diferențele nivelurilor dezvoltării tehnice din diferite ramuri ale muncii din perioada de bază *).

2° Problema alegerii bazei de comparație și a sistemului de ponderi corespunzător la construirea indicilor teritoriali este, după cum am văzut, o problemă cheie. S. Iughenburg a propus metoda coeficienților standard.

Pentru a ajunge la concluzia că numai aplicarea metodei coeficienților standard ar putea rezolva problema alegerii bazei de comparație la construirea indicilor teritoriali, S. Iughenburg a relevat cunoscuta contradicție la care se poate ajunge cînd se ia ca bază fie un obiect, fie altul din colectivitatea de obiecte care trebuie să fie comparate.

Ne permitem să observăm că — în forma prezentată — această contradicție prezintă doar un caz particular și că ea este generală în anumite condiții (condiții ce stau și la baza exemplului folosit de S. Iughenburg).

*) Vezi în această privință considerațiile lui P. Joffe din Vestnik Statistiki, 2, 70—72 (1959).

Scanned with OKEN Scanner

Într-adevăr, o analiză mai adîncă a problemei ne arată acest lucru.
Să considerăm schema teoretică din tabelul 78 pe care ne-a sugerat-o S. Iughenburg.

Tabelul 78

Ramuri ale industriei	Producția globală (q)		Numărul mediu scriptic al muncitorilor (T)		Numărul de om-ani pe o unitate de producție (t)	
	Regiunea A	Regiunea B	Regiunea A	Regiunea B	Regiunea A	Regiunea B
I	q_1	q_1'	T_1	T_1'	t_1	t_1'
II	q_2	q_2'	T_2	T_2'	t_2	t_2'

S-a notat

$$t_i = \frac{T_i}{q_i}, \quad (i = 1,2); \qquad t_i' = \frac{T_i'}{q_i'}, \quad (i = 1,2).$$

Avem

$$I^{\text{ter.}}_{\text{vol. fiz.}(A)} = \frac{\frac{T_1}{q_1} \cdot q_1' + \frac{T_2}{q_2} \cdot q_2'}{T_1 + T_2},$$

$$I^{\text{ter.}}_{\text{vol. fiz.}(B)} = \frac{\frac{T_1'}{q_1'} \cdot q_1 + \frac{T_2'}{q_2'} \cdot q_2}{T_1' + T_2'}.$$

Este evident că dacă ambii indici sînt mai mari decît 1, se ajunge la o contradicție.
Or, ca să avem acest lucru trebuie ca

$$\frac{T_1}{q_1} \cdot q_1' + \frac{T_2}{q_2} \cdot q_2' > T_1 + T_2,$$

$$\frac{T_1'}{q_1'} \cdot q_1 + \frac{T_2'}{q_2'} \cdot q_2 > T_1' + T_2'.$$

Însă

$$T_1 = t_1 q_1; \; T_2 = t_2 q_2; \; T_1' = t_1' q_1'; \; T_2' = t_2' q_2'.$$

Relațiile de mai sus devin:

$$t_1 q_1' + t_2 q_2' > t_1 q_1 + t_2 q_2$$
$$t_1' q_1 + t_2' q_2 > t_1' q_1' + t_2' q_2',$$

sau

$$t_1(q_1 - q_1') < t_2(q_2' - q_2)$$
$$t_1'(q_1' - q_1) < t_2'(q_2 - q_2'),$$

sau

$$t_1(q_1 - q_1') < t_2(q_2' - q_2)$$
$$t_1'(q_1 - q_1') > t_2'(q_2' - q_2).$$

Inegalităţile de mai sus nu pot fi îndeplinite dacă

$$q_1 > q_1', \; q_2 > q_2'.$$

Însă, dacă

$$q_1 - q_1' > 0 \text{ şi } q_2' - q_2 > 0$$

deducem

$$t_2 > \frac{q_1 - q_1'}{q_2' - q_2} t_1,$$

$$t_2' < \frac{q_1 - q_1'}{q_2' - q_2} t_1'.$$

Acestor inegalităţi li se poate da şi o interpretare geometrică din care să rezulte contradicţia de care ne ocupăm şi care se întîmplă adeseori în practică.

După cum am mai precizat, autorul S. Iughenburg dă un exemplu particular, care, evident, trebuie să se încadreze în condiţiile generale pe care le-am examinat mai sus.

Astfel, S. Iughenburg consideră următoarele date:

$$q_1 = 100; \; q_2 = 60; \; T_1 = 16\,000; \; T_2 = 2\,400,$$
$$q_1' = 30; \; q_2' = 400; \; T_1' = 6\,000; \; T_2' = 12\,000.$$

Rezultă:

$$t_1 = 160; \; t_2 = 40; \; t_1' = 200; \; t_2' = 30.$$

Avem

$$q_1 - q_1' = 70 > 0; \; q_2' - q_2 = 340 > 0.$$

Reprezentînd o dreaptă D care are drept ecuaţie:

$$70\,t_1 - 340\,t_2 = 0 \text{ sau } 7\,t_1 - 34\,t_2 = 0$$

vom obţine graficul din figura 41 în plan cu ajutorul axelor perpendiculare Ot_1 şi Ot_2.

Toate punctele $A\,(t_1, t_2)$ din regiunea haşurată prin linii orizontale fac expresia noastră pozitivă şi, invers, toate punctele $A'(t_1', t_2')$, din regiunea haşurată prin linii verticale, fac această expresie negativă. În cazul cînd punctul $A(t_1, t_2) = A\,(160, 40)$ cade deasupra dreptei D (regiunea haşurată prin linii verticale, în care expresia noastră este negativă), iar punctul $A'\,(t_1', t_2') = A'(200, 30)$ — dedesubtul dreptei D (regiunea haşurată prin linii orizontale, în care expresia noastră este pozitivă), se ajunge la contradicţia remarcată de S. Iughenburg.

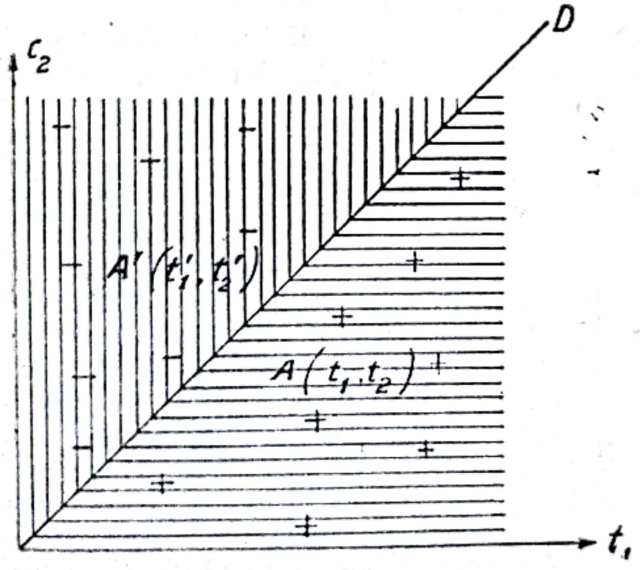


Fig. 41

Folosind procedeul de reprezentare de mai sus, putem găsi nenumărate exemple care duc la aceeaşi contradicţie.

Pentru a ieşi din contradicţia relevată, S. Iughenburg — ca şi alţi autori — consideră

că la construirea indicilor teritoriali se poate aplica cu succes metoda coeficienților standard.

Și în legătură cu această propunere observăm că, în esența ei, această metodă are un mare neajuns, și anume acela că estompează specificul unității comparate căreia i se atribuie o anumită pondere („pondere standard") pentru a-i compara caracteristica studiată cu o altă caracteristică, de aceeași natură, a unei alte unități. Se pot da numeroase exemple cînd, eliminîndu-se structura efectivă a colectivității studiate și înlocuind-o cu o altă structură „standard" generală, se măresc artificial nivelurile medii ale caracteristicii în acele grupe a căror greutate specifică în ansamblul colectivității comparate este mică și, invers, se micșorează artificial nivelurile medii ale aceleiași caracteristici cînd greutatea ei specifică în colectivitatea comparată este mare, iar în colectivitatea „standard" — mai mică.

Pentru a determina volumul producției din cele două regiuni din exemplul luat, S. Iughenburg folosește drept pondere standard nivelul mediu al cheltuielilor de timp de muncă vie necesare pentru a crea un produs, calculat pe ramuri și pe ambele regiuni; această pondere se combină cu volumul producției efective (q_1) pe ramuri, obținîndu-se astfel doi indicatori ai volumului producției pentru fiecare din cele două regiuni, exprimați în om-ani, care devin comparabili, eliminîndu-se totodată și contradicția de care ne-am ocupat.

De fapt, procedeul folosit de S. Iughenburg nu este acela al „standardizării structurii" ci acela al standardizării unei caracteristici calitative, structura colectivității rămînînd cea efectivă. Este, după cum știm, așa-numitul procedeu indirect al metodei coeficienților standard.

3⁰. Este necesară găsirea unui sistem de indici teritoriali corelați, în sensul sistemelor generale de corelare a indicilor, folosite în statistica socialistă.

Acest lucru îl obține L. M. Satunovski, P. Joffe și alții.

Pentru a releva conținutul și eficacitatea analitică a sistemului propus de L. M. Satunovski, îl vom aplica la exemplul folosit de S. Iughenburg (vezi tabelul 72).

Putem calcula următorii indici teritoriali:

Pentru regiunea A

a) $$I^{\text{ter.}}_{\text{vol. fiz.}_{(A)}} = \frac{\Sigma \bar{t} q_{(A)}}{\Sigma \bar{t} Q} =$$

$$= \frac{169{,}2 \cdot 100 + 31{,}3 \cdot 60}{169{,}2 \cdot 130 + 31{,}3 \cdot 460} =$$

$$= \frac{18\,798}{36\,394} = 0{,}5165$$

Pentru regiunea B

a) $$I^{\text{ter.}}_{\text{vol. fiz.}_{(B)}} = \frac{\Sigma \bar{t} q_{(B)}}{\Sigma \bar{t} Q} =$$

$$= \frac{169{,}2 \cdot 30 + 31{,}3 \cdot 400}{169{,}2 \cdot 130 + 31{,}3 \cdot 460} =$$

$$= \frac{17\,596}{36\,394} = 0{,}4835$$

b) $I^{ter.}_{pr.\,m.(A)} = \dfrac{\Sigma \bar{t} q_{(A)}}{\Sigma \bar{t}_{(A)} q_{(A)}} =$

$= \dfrac{169{,}2 \cdot 100 + 31{,}3 \cdot 60}{160 \cdot 100 + 40 \cdot 60} =$

$= \dfrac{18\,798}{18\,400} = 1{,}0216$

c) $I^{ter.}_{ch.\,timp.(A)} = \dfrac{\Sigma t_{(A)} q_{(A)}}{\Sigma \bar{t} Q} =$

$= \dfrac{18\,400}{36\,394} = 0{,}5056$

b) $I^{ter.}_{pr.\,m.(B)} = \dfrac{\Sigma \bar{t} q_{(B)}}{\Sigma t_{(B)} q_{(B)}} =$

$= \dfrac{17\,596}{200 \cdot 30 + 30 \cdot 400} =$

$= \dfrac{17\,596}{18\,000} = 0{,}9776$

c) $I^{ter.}_{ch.\,timp.(B)} = \dfrac{\Sigma t_{(B)} q_{(B)}}{\Sigma \bar{t} Q} =$

$= \dfrac{18\,000}{36\,394} = 0{,}4946$

Sistemele se verifică :

$I^{ter.}_{vol.\,fiz.(A)} = I^{ter.}_{pr.\,m.(A)} \cdot I^{ter.}_{ch.timp.(A)}$

$0{,}5165 = 1{,}0216 \cdot 0{,}5056$

$I^{ter.}_{vol.\,fiz.(B)} = I^{ter.}_{pr.\,m.(B)} \cdot I^{ter.}_{ch.\,timp.(B)}$

$0{,}4835 = 0{,}9776 \cdot 0{,}4946$

O interesantă remarcă se poate face în legătură cu această aplicație. Se observă că dacă comparăm indicele teritorial al volumului fizic al activității de producție calculat pentru regiunea A (0,5165) cu indicele volumului fizic al activității de producție calculat pentru regiunea B (0,4835), obținem un indicator care este identic cu acela pe care l-a obținut și S. Iughenburg. Într-adevăr,

$$I^{ter.}_{vol.\,fiz.}\left(\frac{A}{B}\right) = \frac{\Sigma \bar{t} q_{(A)}}{\Sigma \bar{t} Q} : \frac{\Sigma \bar{t} q_{(B)}}{\Sigma \bar{t} Q} = \frac{\Sigma \bar{t} Q_{(A)}}{\Sigma \bar{t} Q_{(B)}} .$$

În cazul datelor concrete,

$$I^{ter.}_{vol.\,fiz.}\left(\frac{A}{B}\right) = \frac{0{,}5165}{0{,}4835} = 1{,}068 \text{ (vezi (XVI, 3.1))}$$

Considerăm că nici nu se putea obține un alt rezultat, întrucît și în sistemul propus de L. M. Satunovski se iau anumite caracteristici ale colectivității generale drept ponderi la construirea indicilor teritoriali; este vorba, deci, tot de un procedeu de „standardizare" a indicilor teritoriali, pentru a găsi criterii juste, obiective, de comparație.

Avantajul sistemului folosit de L. M. Satunovski constă în faptul că aplică principiile generale de ponderare consacrate în statistica socialistă, pentru a oglindi comparativ (dinamic sau static) latura cantitativă, în legătură indisolubilă cu latura calitativă, a fenomenelor și proceselor vieții social-economice. De asemenea, el oferă posibilitatea de a fi comparate caracteristicile (calitative și cantitative) ale fiecărei regiuni în

Scanned with OKEN Scanner

parte, cu caracteristicile sintetice ale întregii colectivități studiate, precum și cu caracteristicile fiecărei regiuni în parte, una față de alta.

4°. Am atras atenția că formula de bază a indicelui teritorial al productivității muncii propusă de P. Joffe este identică cu formula lui L. M. Satunovski.

Se observă totuși că pentru aplicații practice este mai potrivită formula lui P. Joffe.

Indicatorul K din această formulă este mai potrivit prin faptul că exprimă nivelul relativ al productivității muncii din întreprinderea respectivă ceea ce-i asigură o valoare autonomă.

O particularitate remarcabilă a indicatorului propus de P. Joffe constă în aceea că, avînd forma mediei aritmetice, este mai convenabil la efectuarea calculelor, înlesnindu-le considerabil. Pe de altă parte, ponderile sînt exprimate în cantități de timp de muncă (sau număr de muncitori) ceea ce iarăși înlesnește, standardizează calculele necesare analizei statistico-economice.

5°. În funcție de condițiile concrete existente în întreprinderile din două sau mai multe regiuni comparate ca și în funcție de datele existente, este foarte utilă folosirea indicelui teritorial sintetic al productivității muncii, eliberat de influența deosebirilor în structură (XVI, 3.22), identic cu indicele lui L. M. Satunovski de forma arătată.

6°. Am văzut că sistemul de indici teritoriali propus de L. Kazineț se fundamentează, în esență, pe metoda coeficienților standard care-l apropie, din punct de vedere metodologic, de sistemul grupului S. Iughenburg, G. Baklanov și Z. Sidlik.

Pe de altă parte, am mai văzut că aplicînd sistemul propus de L. M. Satunovski la exemplul folosit de S. Iughenburg, metodele de calcul propuse de acești autori duc la rezultate identice.

Acest lucru a fost posibil tocmai datorită faptului că și în sistemul propus de L. M. Satunovski, drept ponderi la construirea indicilor teritoriali se iau anumite caracteristici ale întregii colectivități de obiecte comparate (întreprinderi, regiuni etc.), ceea ce înseamnă că este vorba tot de un procedeu de „standardizare" a ponderilor.

În mod firesc, se impune să examinăm problema dacă nu s-ar putea stabili o verigă de legătură și între sistemul lui L. Kazineț și cel propus de L. M. Satunovski.

Socotim că într-adevăr această legătură se poate stabili, dacă comparăm cele două sisteme sub aspectul conținutului lor.

Am văzut că sistemul lui L. M. Satunovski oferă posibilitatea de a fi comparate caracteristicile calitative și cantitative ale fiecărei regiuni în parte cu caracteristicile sintetice ale întregii colectivități studiate, *dar și cu caracteristicile fiecărei regiuni în parte, una față de alta.*

Să concretizăm această din urmă posibilitate împărțind cei trei indici corelați ai lui L. M. Satunovski pentru o întreprindere, la indicii corespunzători ai altei întreprinderi.

Scanned with OKEN Scanner

Obţinem astfel :

Indicele volumului fizic al activităţii de producţie

$$\frac{\Sigma \bar{t} q_1}{\Sigma \bar{t} Q} : \frac{\Sigma \bar{t} q_2}{\Sigma \bar{t} Q} = \frac{\Sigma \bar{t} q_1}{\Sigma \bar{t} q_2}; \qquad \text{(XVI, 4.1)}$$

Indicele productivităţii muncii

$$\frac{\Sigma \bar{t} q_1}{\Sigma t_1 q_1} : \frac{\Sigma \bar{t} q_2}{\Sigma t_2 q_2}; \qquad \text{(XVI, 4.2)}$$

Indicele cheltuielilor de timp de muncă

$$\frac{\Sigma t_1 q_1}{\Sigma \bar{t} Q} : \frac{\Sigma t_2 q_2}{\Sigma \bar{t} Q} = \frac{\Sigma t_1 q_1}{\Sigma t_2 q_2}. \qquad \text{(XVI, 4.3)}$$

Acest din urmă indice se poate obţine folosind şi legătura de sistem dintre indicele caracteristicii cantitative (XVI, 4.1) şi cel privind caracteteristica calitativă (XVI, 4.2). Într-adevăr, făcînd raportul dintre formulele (XVI, 4.1) şi (XVI, 4.2) obţinem

$$\frac{\Sigma \bar{t} q_1}{\Sigma \bar{t} q_2} : \left(\frac{\Sigma \bar{t} q_1}{\Sigma t_1 q_1} : \frac{\Sigma \bar{t} q_2}{\Sigma t_2 q_2} \right) = \frac{\Sigma t_1 q_1}{\Sigma t_2 q_2}. \qquad \text{(XVI, 4.3a)}$$

Or, dacă comparăm relaţiile (XVI, 4.3) şi (XVI, 4.3a) cu relaţiile (XVI, 3.15) şi (XVI, 3.16) din sistemul lui L. Kazineţ, constatăm o deplină identitate (forma diferită a notaţiei nu poate afecta identitatea de conţinut).

7°. Sistemele pe care le-am examinat (cu excepţia sistemului lui V. Peregudov) confirmă şi mai mult teza pe care am demonstrat-o — în condiţiile cele mai generale, vezi punctul 2° de mai sus — că la elaborarea indicilor teritoriali, respectînd schema de ponderare acceptată în general în teoria indicilor, nici una din regiunile, întreprinderile etc. comparate, nu poate fi luată drept bază de comparaţie.

8°. Cu privire la sistemul indicilor teritoriali propus de V. Peregudov, remarcăm următoarele :

Ca şi toţi ceilalţi autori, V. Peregudov constată că sarcina comparării pe regiuni a indicatorilor economici calitativi nu se poate realiza prin raportări directe, cu folosirea sistemului de ponderare adoptat în general. Comparări directe s-ar putea face numai în cazul cînd ar fi satisfăcută cerinţa „reversibilităţii" şi în cazul indicilor agregaţi.

Examinînd mai multe exemple, V. Peregudov ajunge la concluzia că „reversibilitatea" indicilor agregaţi este asigurată numai dacă între

variația indicatorilor individuali care intră în compunerea indicelui de grup (am văzut că în studiul lui V. Peregudov este vorba de indicii individuali ai prețului de cost și ai volumului de producție pe feluri de produse, notați respectiv prin i_c și i_q), nu există nici un fel de corelație. Cu alte cuvinte, condiția cea mai generală a „reversibilității" indicilor agregați constă în independența sistemului indicilor i_c față de sistemul indicilor i_q.

Or — arată V. Peregudov — criteriul independenței a două mărimi x și y este dat de egalitatea

$$\overline{x \cdot y} = \bar{x} \cdot \bar{y}.$$

Pe acest principiu teoretic V. Peregudov își fundamentează apoi sistemul de indici dat de formulele (XVI, 3.25).

V. Peregudov nu demonstrează însă fondul teoretic al criteriului de independență propus, ceea ce face dificilă înțelegerea întregului sistem pe care l-a conceput.

Să lămurim dificultatea în care ne aflăm:

V. Peregudov demonstrează relația

$$\frac{\Sigma i_c i_q w_1}{\Sigma w_1} = \frac{\Sigma i_c w_1}{\Sigma w_1} \cdot \frac{\Sigma i_q w_1}{\Sigma w_1}.$$

Față de sistemul de ponderi w_1, putem scrie

$$\frac{\Sigma i_c w_1}{\Sigma w_1} = \bar{i}_c; \quad \frac{\Sigma i_q w_1}{\Sigma w_1} = \bar{i}_q; \quad \frac{\Sigma i_c i_q w_1}{\Sigma w_1} = \overline{i_c i_q}.$$

Prin urmare,

$$\overline{i_c i_q} = \bar{i}_c \cdot \bar{i}_q.$$

De aici Peregudov deduce că variabilele i_c și i_q sînt independente, afirmînd textual: „Condiția independenței a două mărimi x și y constă în egalitatea produsului mediu $\overline{xy}$ cu produsul mediilor $\bar{x} \cdot \bar{y}$."

Or, după cum știm din teoria generală a corelației, este adevărat că dacă x și y sînt variabile independente, avem

$$\overline{x \cdot y} = \bar{x} \cdot \bar{y},$$

însă, proprietatea inversă — pe care a invocă Peregudov — nu este adevărată.

Într-adevăr, putem avea numeroase cazuri de variabile x și y care nu sînt independente și să avem totuși îndeplinită relația $\overline{xy} = \bar{x} \cdot \bar{y}$.

Scanned with OKEN Scanner

Acest lucru poate fi scos în evidență și prin următorul exemplu: Să considerăm tabloul de corelație a două variabile x și y (tabelul 79).

Tabelul 79

x \ y	-1	0	1	Σ
0	$\frac{4}{20}$	$\frac{2}{20}$	$\frac{4}{20}$	$\frac{10}{20}$
1	$\frac{3}{20}$	$\frac{4}{20}$	$\frac{3}{20}$	$\frac{10}{20}$
Σ	$\frac{7}{20}$	$\frac{6}{20}$	$\frac{7}{20}$	1

Variabila x ia valorile 0 și 1 cu probabilitățile $\frac{10}{20}$, $\frac{10}{20}$ iar variabila y ia valorile -1, 0, 1, cu probabilitățile respective: $\frac{7}{20}, \frac{6}{20}, \frac{7}{20}$.

Avem

$$\bar{x} = \frac{10}{20} \cdot 0 + \frac{10}{20} \cdot 1 = \frac{1}{2},$$

$$\bar{y} = \frac{7}{10} \cdot (-1) + \frac{6}{20} \cdot 0 + \frac{7}{20} \cdot 1 = 0,$$

$$\overline{x \cdot y} = -\frac{3}{20} + \frac{3}{20} = 0.$$

Prin urmare

$$\overline{xy} = 0, \ \bar{x} \cdot \bar{y} = \frac{1}{2} \cdot 0 = 0,$$

$$\overline{x.y} = \bar{x} \cdot \bar{y}.$$

Însă, variabilele x și y nu sînt *independente*.
Într-adevăr, după tabloul de corelație avem probabilitățile:

$$P(x = 0, \ y = -1) = \frac{4}{20},$$

$$P(x = 0) = \frac{10}{20},$$

$$P(y = -1) = \frac{7}{20}.$$

Nu avem însă relaţia

$$P(x = 0,\ y = -1) = P(x = 0) \cdot P(y = -1), \qquad \text{(XVI, 4.4)}$$

deoarece

$$\frac{4}{20} \neq \frac{10}{20} \cdot \frac{7}{20}.$$

Aşadar, variabilele x şi y nu sînt independente, deşi $\overline{x \cdot y} = \bar{x} \cdot \bar{y}$; relaţia (XVI, 4,4) care exprimă condiţia că probabilitatea îndeplinirii a două evenimente independente este egală cu produsul probabilităţilor celor două evenimente în parte, nu se realizează în speţă.

*

Apreciem, în concluzie, că din punct de vedere teoretic, problema elaborării indicilor teritoriali a fost abordată multilateral, deşi se mai impun încă unele precizări de amănunt. Ne referim, de pildă, la necesitatea de a fi precizate, teoretic, sfera şi conţinutul economic al indicatorului producţiei globale calculat după metoda economiei naţionale (indicator propus de G. Baklanov care are o mare importanţă practică) sau unele condiţii suplimentare în fundamentarea teoretică a indicilor teritoriali propuşi de V. Peregudov.

Scanned with OKEN Scanner

Scanned with OKEN Scanner

Capitolul XVII

Aplicaţii ale coeficientului de corelaţie liniară în lucrări de analiză prin metoda indicilor

1. Consideraţii introductive

Am văzut în capitolul XI că în cercetarea fenomenelor social-economice, coeficientul de corelaţie liniară poate servi drept instrument de analiză deosebit de preţios şi că, în general, el poate fi folosit cu succes la rezolvarea multor probleme ale teoriei şi practicii ştiinţei statistice.

În acest capitol vom încerca să arătăm că funcţiile analitice ale coeficientului de corelaţie liniară se vădesc tot atît de bune şi la analiza sau interpretarea rezultatelor obţinute prin metoda indicilor, avînd convingerea că abordarea şi sub acest aspect a problemelor pe care le ridică calculul indicilor prezintă un interes teoretic şi practic incontestabil.

În capitolul XV am văzut că mărimea unui indice agregat se schimbă odată cu schimbarea sistemului de ponderare. Astfel, indicele agregat al preţurilor la construirea căruia am folosit drept ponderi structura indicatorului cantitativ din perioada curentă, diferă sensibil de indicele analog construit pe baza ponderilor din perioada de bază. Diferă substanţial şi indicii productivităţii muncii calculaţi prin metode diferite.

Se pune de aceea, în mod firesc, problema de a şti care este corelaţia dintre doi indici agregaţi, la al căror calcul s-au folosit sisteme diferite de ponderare, de a releva şi interpreta diferenţa de mărime constatată la aceşti indici ca şi factorii care au determinat-o.

Este evident că un interes deosebit îl prezintă şi stabilirea condiţiilor celor mai generale, necesare şi suficiente, în care putem avea egalitatea între doi indici agregaţi în construcţia cărora au intrat ponderi diferite.

Alegînd exemplul indicilor agregaţi nu ne-am gîndit, desigur, să limităm la acest cadru aplicarea coeficientului de corelaţie liniară — după cum vom vedea mai jos. El se poate aplica de asemenea la analiza core-

laţiei dintre indicele cu structură variabilă de tipul] $\bar{\alpha} = \frac{\bar{\alpha}_1}{\bar{\alpha}_0}$ şi indicele dinamicii medii ca şi la analiza corelaţiei dintre alţi indici (corelaţia dintre indicele în expresie valorică şi indicele în unităţi de timp de muncă etc.).

De aceste probleme interesante s-a ocupat într-un studiu amplu, cunoscutul statistician sovietic L. S. Kazineţ *), care a analizat posibilităţile aplicării eficiente a coeficientului de corelaţie liniară în domeniul indicilor.

2. Corelaţia dintre două mărimi medii calculate pe baza ponderilor diferite

Să considerăm o caracteristică variabilă x şi să-i calculăm două valori medii cu ponderi diferite M şi N. Vom avea deci:

$$\bar{x}_M = \frac{\Sigma xM}{\Sigma M} \quad \text{şi} \quad \bar{x}_N = \frac{\Sigma xN}{\Sigma N}.$$

Făcînd diferenţa dintre aceste valori medii şi împărţind numărătorul şi numitorul primei medii prin numitorul celei de-a doua, obţinem

$$\bar{x}_M - \bar{x}_N = \frac{\Sigma xM}{\Sigma M} - \frac{\Sigma xN}{\Sigma N} = \frac{\frac{\Sigma xM}{\Sigma N}}{\frac{\Sigma M}{\Sigma N}} - \frac{\Sigma xN}{\Sigma N}.$$

Aducînd la acelaşi numitor diferenţa de mai sus şi înlocuind pe M cu $\frac{M}{N} \cdot N$, putem scrie:

$$\bar{x}_M - \bar{x}_N = \frac{\frac{\Sigma x \frac{M}{N} N}{\Sigma N} - \frac{\Sigma xN}{\Sigma N} \cdot \frac{\Sigma \frac{M}{N} N}{\Sigma N}}{\frac{\Sigma \frac{M}{N} N}{\Sigma N}} = \frac{\overline{x \cdot \frac{M}{N}} - \bar{x} \cdot \left(\overline{\frac{M}{N}}\right)}{\left(\overline{\frac{M}{N}}\right)},$$

în care termenii cu bară sînt mărimi medii.

Dacă notăm $y = \frac{M}{N}$, observăm că la numărătorul raportului obţinut avem diferenţa dintre produsul mediu a două variabile şi produsul mediilor lor.

*) L. S. Kazineţ, *Koefiţient koreliaţii i voprosi teorii indeksov*. Ucinolie zapiski po statistike, I, 125–138.

Cum din formula coeficientului de corelație

$$\rho_{xy} = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{\sigma_x \sigma_y},$$

deducem că

$$\overline{xy} - \bar{x} \cdot \bar{y} = \rho_{xy} \cdot \sigma_x \cdot \sigma_y;$$

în cazul nostru vom avea definitiv

$$\bar{x}_M - \bar{x}_N = \frac{\Sigma xM}{\Sigma M} - \frac{\Sigma xN}{\Sigma N} = \frac{\rho_{x\frac{M}{N}} \cdot \sigma_x \cdot \sigma_{\frac{M}{N}}}{\left(\frac{\overline{M}}{N}\right)} = \rho_{x\frac{M}{N}} \sigma_x \cdot V_{\frac{M}{N}}, \qquad \text{(XVII, 2.1)}$$

în care:

$\rho_{x\frac{M}{N}}$ este coeficientul de corelație între valorile caracteristicii variabile x și rapoartele ponderilor $\frac{M}{N}$ ale celor două medii considerate;

σ_x, abaterea medie pătratică a caracteristicii variabile x;

$V_{\frac{M}{N}}$, coeficientul de variație a rapoartelor $\frac{M}{N}$ dintre ponderi.

Procedînd analog, obținem și următoarea expresie a raportului geometric dintre cele două medii:

$$\frac{\Sigma xM}{\Sigma M} : \frac{\Sigma xN}{\Sigma N} = 1 + \rho_{x\frac{M}{N}} \cdot V_x \cdot V_{x\frac{M}{N}} \qquad \text{(XVII, 2.1a)}$$

în care V_x este coeficientul de variație a caracteristicii x.

Expresia (XVII,2.1 a) ne arată că raportul dintre două medii calculate pe bază de ponderi diferite se exprimă simplu în funcție de trei variabile:

1°. Coeficientul de corelație dintre valorile caracteristicii variabile x și rapoartele corespunzătoare dintre ponderile diferite a celor două medii ale variabilei x;

2°. Coeficientul de variație a caracteristicii x;

3°. Coeficientul de variație a rapoartelor dintre ponderile $\frac{M}{N}$.

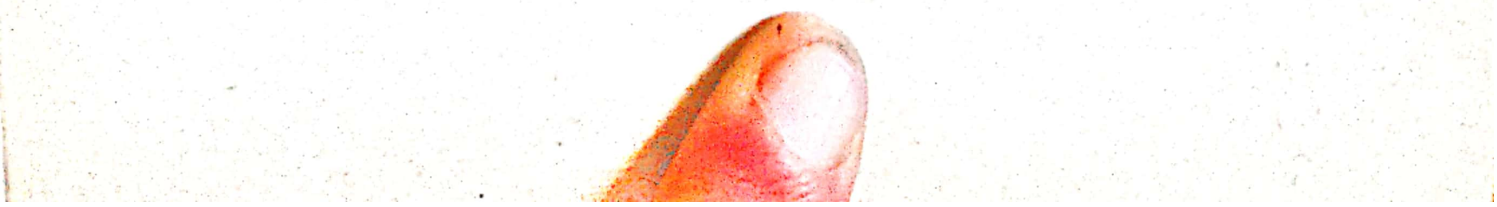
Din formula (XVII,2.1) se vede că

$$\frac{\Sigma xM}{\Sigma M} = \frac{\Sigma xN}{\Sigma N}$$

dacă se realizează una din egalitățile:

$$\rho_{x\frac{M}{N}} = 0, \quad \sigma_x = 0 \text{ sau } V_{\frac{M}{N}} = 0.$$

Egalitatea $\rho_{x\frac{M}{N}} = 0$ este o condiție suficientă ca să aibă loc egalitatea dintre cele două medii.



Cu alte cuvinte, *lipsa unei corelații liniare dintre valorile caracteristicii* x *și rapoartele* $\frac{M}{N}$, *asigură egalitatea*

$$\bar{x}_M = \bar{x}_N .$$

Aceeași egalitate dintre medii este asigurată și în cazul cînd V_x sau $V_{\frac{M}{N}}$ sînt egale cu zero.

Cazul $V_x = 0$ are loc dacă toate valorile caracteristicii x sînt egale între ele: $x_1 = x_2 = x_3 \ldots$; se vede că este un caz banal în care $\bar{x} = x_i$ și în care avem deci o independență deplină dintre valoarea medie $\bar{x}$ și sistemul respectiv de ponderi.

Cazul $V_{\frac{M}{N}} = 0$ are loc cînd cele două sisteme de ponderi sînt proporționale, adică, cînd avem egalitățile:

$$\frac{M_1}{N_1} = \frac{M_2}{N_2} = \ldots = \frac{M_i}{N_i} = k \text{ (constantă)}.$$

Din cele de mai sus rezultă, însă, că cea mai generală condiție — necesară și suficientă — pentru ca cele două medii aritmetice $\bar{x}_M$ și $\bar{x}_N$ să fie egale este ca numărătorul coeficientului de corelație să fie egal cu zero, adică să avem egalitatea

$$\overline{x \cdot \frac{M}{N}} - \bar{x} \cdot \left(\overline{\frac{M}{N}}\right) = 0, \qquad \text{(XVII, 2.2)}$$

care este echivalentă cu egalitatea

$$\overline{x \cdot \frac{M}{N}} = \bar{x} \cdot \left(\overline{\frac{M}{N}}\right).$$

Cele trei condiții examinate mai sus nu sînt decît cazuri particulare ale acestei condiții generale; este evident că odată îndeplinită condiția (XVII,2.2), are loc și una din egalitățile

$$\rho_{x\frac{M}{N}} = 0, \quad V_x = 0, \quad V_{\frac{M}{N}} = 0.$$

În toate celelalte cazuri, două medii calculate pe baza ponderilor diferite $\bar{x}_M$ și $\bar{x}_N$ nu vor fi egale, iar diferența dintre ele va fi determinată de valorile celor trei elemente componente ale formulei (XVII, 2.1). Semnul acestei diferențe va fi dat de semnul coeficientului de corelație liniară dintre valorile caracteristicii x și rapoartele $\frac{M}{N}$ ale ponderilor.

Scanned with OKEN Scanner

Vom avea inegalitățile:

Dacă $\rho_{x\frac{M}{N}} > 0$, atunci $\bar{x}_M > \bar{x}_N$.

Dacă $\rho_{x\frac{M}{N}} < 0$, atunci $\bar{x}_M < \bar{x}_N$.

Din formulele (XVII, 2.1) și (XVII, 2.1 a) pot fi deduse ușor formulele care exprimă corelația dintre media aritmetică ponderată și media aritmetică simplă.

Într-adevăr, presupunînd că în formulele (XVII, 2.1) și (XVII, 2.1a) toate ponderile celei de-a doua medii $\bar{x}_N$ sînt egale cu 1 ($N_i = 1$), vom avea egalitățile

$$\frac{\Sigma xM}{\Sigma M} - \frac{\Sigma x}{s} = \rho_{xM} \cdot \rho_x \cdot V_M \qquad \text{(XVII, 2.3)}$$

și

$$\frac{\Sigma xM}{\Sigma M} : \frac{\Sigma x}{s} = 1 + \rho_{xM}\, V_x\, V_M \qquad \text{(XVII, 2.3a)}$$

în care s este numărul variantelor caracteristicii x.

3. Corelația dintre indicii agregați cu ponderi diferite. Indici ai prețurilor

Să considerăm, de exemplu, doi indici agregați ai prețurilor de tipul B (XV, 1.4) și (XV, 1.5):

$$I_{p(q_1)} = \frac{\Sigma p_1 q_1}{\Sigma p_0 q_1} \quad \text{și} \quad I_{p(q_0)} = \frac{\Sigma p_1 q_0}{\Sigma p_0 q_0}.$$

Diferența δ dintre acești doi indici este

$$\delta = \frac{\Sigma p_1 q_1}{\Sigma p_0 q_1} - \frac{\Sigma p_1 q_0}{\Sigma p_0 q_0}$$

sau, considerînd ambii indici în forma lor de medii aritmetice,

$$\delta = \frac{\Sigma \frac{p_1}{p_0} p_0 q_1}{\Sigma p_0 q_1} - \frac{\Sigma \frac{p_1}{p_0} p_0 q_0}{\Sigma p_0 q_0}.$$

În această formă, compararea a doi indici agregați cu ponderi diferite se reduce la compararea a două medii ponderate, ceea ce ne permite să aplicăm formula (XVII, 2.1), în transformările corespunzătoare ținînd seama că $x = \frac{p_1}{p_0} = i_p$, $M = p_0 q_1$ și $N = p_0 q_0$.

Vom avea deci

$$\delta = \rho_{i_p i_q} \cdot \sigma_{i_p} \cdot V_{\frac{p_0 q_1}{p_0 q_0}} = \rho_{i_p i_q} \cdot \sigma_{i_p} \cdot V_{i_q}, \qquad \text{(XVII, 3.1)*)}$$

în care:

$\rho_{i_p i_q}$ simbolizează coeficientul de corelație dintre indicii individuali ai prețurilor $\left(i_p = \frac{p_1}{p_0}\right)$ și indicii individuali ai ponderilor $\left(i_q = \frac{q_1}{q_0}\right)$;

σ_{i_p}, abaterea medie pătratică a indicilor individuali ai prețurilor;

V_{i_q}, coeficientul de variație a indicilor individuali ai cantităților de produse.

Raportul dintre cei doi indici agregați va căpăta forma:

$$\frac{I_{p(q_1)}}{I_{p(q_0)}} = \frac{\Sigma p_1 q_1}{\Sigma p_0 q_1} : \frac{\Sigma p_1 q_0}{\Sigma p_0 q_0} = 1 + \rho_{i_p i_q} V_{i_p} V_{i_q}, \qquad \text{(XVII, 3.1a)}$$

în care V_{i_p} reprezintă coeficientul de variație a indicilor individuali ai prețurilor.

Dn formula (XVII, 3.1a) rezultă că raportul dintre cei doi indici considerați este determinat de trei elemente componente:

a) de coeficientul de corelație dintre indicii individuali ai prețurilor și cei ai cantităților de mărfuri;

b) de coeficientul de variație a indicilor individuali ai prețurilor și

c) de coeficientul de variație a indicilor individuali ai cantităților.

Rezultă de asemenea că egalitatea

$$I_{p(q_1)} = I_{p(q_0)}$$

este asigurată dacă are loc una din egalitățile:

$$\rho_{i_p i_q} = 0, \quad \sigma_{i_p} = 0, \quad V_{i_p} = 0, \quad V_{i_q} = 0.$$

Să examinăm acum condițiile generale în care au loc următoarele inegalități:

$$I_{p(q_1)} > I_{p(q_0)} \quad \text{și} \quad I_{p(q_1)} < I_{p(q_0)}.$$

Știm că coeficienții de variație V_{i_p} și V_{i_q} sînt întotdeauna mărimi pozitive în timp ce coeficientul de corelație $\rho_{i_p i_q}$ poate fi sau pozitiv sau negativ, în funcție de caracterul corelației liniare.

---

*) Formula (XVII, 3.1) a fost dedusă de W. Bortkievicz în 1923. De aceeași problemă s-au ocupat și următorii autori: B.S. Iastremski, *Sleduet li primeniat v indeksiom analize metodi variaționnoi statistiki*? Referati i annotații dokladov po statistike, Gosplanizdat, 1949, p. 28—31; A. I. Boiarski, *Nekatorîie primeneniia lineinovo koefițienta koreliații v teorii indeksov.* Problemi uceta o statistiki, 1934, vol. I, ediția 1, p. 49—51.

Problema a fost abordată și de N. V. Peregudov în Ucinolie zapiski po statistike, vol. 1, Moskva, 1955, ca și de G. Edelganz în Vestnik Statistiki, *11*, 12—55 (1959) și de alți autori.

În țara noastră s-a ocupat de această problemă L. Tövissi, *Aplicarea coeficientului de corelație în teoria indicilor*, Revista de Statistică, *4*, 52—70 (1961).

Scanned with OKEN Scanner

De aceea dacă între indicii i_p și i_q avem corelația directă, aceasta înseamnă că

$$\rho_{i_p i_q} > 0 \text{ și atunci } I_{p(q_1)} > I_{p(q_0)}$$

și invers, în cazul unei corelații inverse avem

$$\rho_{i_p i_q} < 0 \text{ și atunci } I_{p(q_1)} < I_{p(q_0)}.$$

Se mai pune problema de a estima — în sensul celor arătate mai jos — mărimea relativă și absolută a diferenței dintre indicii considerați.

Mărimea diferenței relative este dată de formula (XVII, 3.1 a). În cazul unei variații mai puțin pronunțate a indicilor individuali i_p și i_q (de exemplu, cînd V_{i_p} și V_{i_q} nu depășesc 30%), diferența dintre cei doi indici agregați poate să atingă aproximativ 9% (0,3 ·0,3 = 0,09). În majoritatea cazurilor însă, mărimea relativă a diferenței trebuie să fie mult mai mică întrucît coeficientul corelației liniare este de obicei mult mai mic de 1, iar produsul a doi coeficienți de variație va fi în puține cazuri egal cu 0,09.

Mărimea diferenței absolute δ, dintre indicii $I_{p(q_1)}$ și $I_{p(q_0)}$ poate fi calculată tot pe baza formulei (XVII, 3.1 a).

Într-adevăr, egalitatea (XVII, 3.1 a) poate fi scrisă astfel:

$$\delta = I_{p(q_1)} - I_{p(q_0)} = I_{p(q_0)} : \quad \rho_{i_p i_q} V_{i_p} V_{i_q}, \qquad \text{(XVII, 3.2)}$$

de unde rezultă că putem avea trei situații:

a) $I_{p(q_0)} < 1$, în care caz mărimea absolută a diferenței este mai mică decît mărimea relativă a diferenței;

b) $I_{p(q_0)} > 1$, mărimea absolută a diferenței este mai mare decît cea relativă și

c) $I_{p(q_0)} = 1$, mărimea absolută și relativă a diferenței au aceeași valoare numerică.

Interesul pratic pe care-l prezintă formulele (XVII, 3.1) și (XVII, 3.1a) este incontestabil. După cum arată I. Kazineț, din ele se deduce, de pildă, cunoscuta constatare că indicele prețurilor cu amănuntul, construit pe baza ponderilor din perioada curentă, relativ la cantitățile de mărfuri desfăcute în perioada curentă (q_1), este întotdeauna mai mic decît același indice construit însă pe baza ponderilor din perioada de bază (q_0). Această constatare practică își găsește explicația tocmai în existența unei corelații inverse dintre indicii individuali ai prețurilor și cei ai cantităților de mărfuri desfăcute. Se știe doar că volumul consumului diferitelor mărfuri crește o dată cu reducerea prețurilor.

Corelația inversă se mai constată în practică și între gradul de îndeplinire a planului de reducere a prețului de cost și gradul de îndeplinire a programului de producție a diferitelor produse. Existența acestei corelații face ca indicele de îndeplinire a planului de reducere a prețului de cost, ponderat cu cantitățile efectiv produse (q_i), să fie mai mic decît același indice ponderat cu cantitățile ce s-au planificat a fi produse (q_{pl}).

Indici ai productivității muncii (valoric în unități de timp de muncă). Formula (XVII, 3.1) ne oferă posibilitatea de a releva corelația dintre indi-

cele productivității muncii I_w, calculat în expresie bănească (metoda producției globale medii ce revine pe o unitate de timp de muncă) și indicele productivității muncii calculat în unități de timp de muncă I_t.

După cum știm, indicele productivității muncii calculat după metoda producției globale se calculează după formula (XV, 2.39):

$$I_w = \frac{\Sigma p q_1}{\Sigma t_1 q_1} : \frac{\Sigma p q_0}{\Sigma t_0 q_0}$$

în care la numărătorul fiecărui raport avem respectiv producția globală din perioada curentă și din perioada de bază, în prețuri constante, iar la numitor, cheltuielile de timp de muncă în om-ore ($\Sigma t_1 q_1 = \Sigma T_1$; $\Sigma t_0 q_0 = \Sigma T_0$).

Pe de altă parte, indicele productivității muncii exprimat în unități de timp de muncă I_t, este dat de cunoscuta formulă

$$I_t = \frac{\Sigma t_0 q_1}{\Sigma t_1 q_1}.$$

Dacă luăm raportul $\frac{I_w}{I_t}$, facem simplificările cuvenite și apoi aplicăm formula (XVII, 3.1a), putem scrie:

$$\frac{I_w}{I_t} = \frac{\frac{\Sigma p q_1}{\Sigma t_1 q_1} : \frac{\Sigma p q_0}{\Sigma t_0 q_0}}{\frac{\Sigma t_0 q_1}{\Sigma t_1 q_1}} = \frac{\Sigma p q_1}{\Sigma p q_0} : \frac{\Sigma t_0 q_1}{\Sigma t_0 q_0} = 1 + \rho_{\frac{p}{t_0} : \frac{q_1}{q_0}} V_{\frac{p}{t_0}} V_{\frac{q_1}{q_0}}. \quad \text{(XVII, 3.3)}$$

Nu e greu de văzut că prima condiție care face ca cei doi indici să fie egali este independența liniară dintre $\frac{p}{t_0}$ (consumul specific de timp de muncă din perioada de bază, exprimat în unități bănești) și $\frac{q_1}{q_0}$ (indicii individuali ai produselor respective în forma lor naturală); mai sînt apoi și alte două condiții:

— proporționalitatea directă între prețurile constante și consumul specific de muncă pentru fiecare produs în parte:

$$\frac{p'}{t_0'} = \frac{p''}{t_0''} = \ldots = \frac{p^i}{t_0^i};$$

— egalitatea rapoartelor

$$\frac{q_1'}{q_0'} = \frac{q_1''}{q_0''} = \ldots = \frac{q_1^i}{q_0^i}.$$

Formula (XVII, 3.3) ne arată și factorii care determină diferența de mărime dintre cei doi indici :

1°. gradul de variație de la un produs la altul, a consumului specific de timp de muncă din perioada de bază, exprimat în unități bănești $\frac{p}{t_0}$;

2°. gradul de variație a indicilor individuali $\frac{q_1}{q_0}$ ai diferitelor produse în forma lor naturală;

3°. gradul de intensitate a legăturii liniare dintre rapoartele $\frac{p}{t_0}$ și $\frac{q_1}{q_0}$.

Ultimii doi factori sînt adeseori trecuți cu vederea în procesul de analiză, considerîndu-se că diferența dintre indicii I_u și I_t s-ar forma exclusiv datorită dispersiei mari a rapoartelor $\frac{p}{t_0}$, ceea ce, evident, constituie o eroare.

Prin urmare, practic, indicele productivității muncii în expresie bănească va fi mai mare decît indicele productivității muncii exprimat în unități de timp de muncă în acele cazuri cînd se intensifică producția acelor produse pentru care, în perioada de bază, a fost atins un înalt nivel — în expresie bănească — al productivității orare a muncii (aceasta se explică prin faptul că în acest caz, rapoartele $\frac{p}{t_0}$ sînt direct proporționale cu rapoartele $i_q = \frac{q_1}{q_0}$). Situația inversă — indicele productivității muncii în expresie bănească va fi mai mic dacă în perioada curentă crește în ritm intens producția acelor produse pentru care în perioada de bază productivitatea orară a muncii în expresie bănească a fost mai mică.

Dat fiind că în activitatea practică de construire a socialismului se planifică și se fabrică produse, în ritm sporit, chiar și acele pentru care productivitatea orară a muncii în perioada de bază era ridicată — aceasta în interesul economiei naționale luate în ansamblu — este evident că vom întîlni cazuri cînd $I_w > I_t$, dar și cazuri cînd $I_w < I_t$.

Interpretarea indicelui deplasărilor structurale cu ajutorul coeficientului de corelație. O funcție cognitivă deosebită îndeplinesc formulele (XVII, 3.1) și (XVII, 3.1a), cu prilejul analizei unor probleme speciale ale teoriei indicilor cum ar fi problema controversată a indicelui deplasărilor structurale*).

După cum am arătat în capitolul XV, prin acest indice se înțelege în mod obișnuit acel indice care se obține prin împărțirea indicelui agregat

*) Este vorba, de fapt, de schimbarea asortimentului și de influența ei asupra indicelui prețului de cost.

ponderat cu ponderile perioadei curente, la indicele agregat ponderat cu ponderile perioadei de bază (vezi formula (XV, 3.5)).

Să revenim, de pildă, la „indicele deplasărilor structurale" (schimbării asortimentului) propus de D. Savinski, folosit în studiul dinamicii prețului de cost:

$$K = \frac{\Sigma z_1 q_1}{\Sigma z_0 q_1} : \frac{\Sigma z_1 q_0}{\Sigma z_0 q_0}.$$

Aplicînd formula (XVII, 3.1a) constatăm că acest „indice al deplasării structurii" este egal cu

$$1 + \rho_{i_z i_q} V_{i_z} V_{i_q}.$$

Or, se vede că mărimea unui asemenea indice este departe de a fi numai o funcție de schimbările care au loc în asortimentul producției de la o perioadă la alta, întrucît aceasta din urmă este caracterizată numai de elementul V_{i_q}. Cu alte cuvinte — în cazul cînd nivelul prețului de cost al diferitelor produse rămîne constant — schimbări de structură pronunțate pot uneori determina schimbări mai mici ale prețului de cost decît acele pe care le determină schimbările de structură neînsemnate. Mai mult, se pot ivi cazuri în care, oricît de mari ar fi schimbările asortimentului producției, ele nu exercită nici o influență asupra indicelui prețului de cost, dacă

$$\rho_{i_z i_q} = 0.$$

În această ultimă ipoteză vom avea

$$I_{z(q_1)} = I_{z(q_0)} \text{ și evident } I_{z(q_1)} : I_{z(q_0)} = 1.$$

L. Kazineț dă exemplul din tabelul 80 pentru a ilustra acest caz.

Tabelul 80

Produse	Prețul de cost unitar		Producția în mii unități	
	de plan	efectiv	planificată	efectivă
A	10	9,5	200	400
B	5	4,5	1 000	200
C	4	2,8	750	750

Aplicînd cunoscutele formule ale indicelui de îndeplinire a sarcinilor de reducere a prețului de cost (vezi formulele XV, 2.2a și XV, 2.2b) la datele din acest exemplu, obținem:

$$I_{z(q_{pl})} = \frac{\Sigma z_1 q_{pl}}{\Sigma z_l q_{pl}} = \frac{(9{,}5 \cdot 200) + (4{,}5 \cdot 1\,000) + (2{,}8 \cdot 750)}{(10 \cdot 200) + (5 \cdot 1\,000) + (4 \cdot 750)} =$$

$$= \frac{8\,500}{10\,000} = 0{,}85 \text{ sau } 85\%$$

și

$$I_{z(q_l)} = \frac{\Sigma z_l q_l}{\Sigma z_{pl} q_l} = \frac{(9{,}5 \cdot 400) + (4{,}5 \cdot 200) + (2{,}8 \cdot 750)}{(10 \cdot 400) + (5 \cdot 200) + (4 \cdot 750)} =$$

$$= \frac{6\,800}{8\,000} = 0{,}85 \text{ sau } 85\%.$$

Se vede clar că deși avem considerabile schimbări în asortimentul producției (vezi datele de plan și cele efective din tabelul 80) ca și reduceri ale prețului de cost la fiecare produs, totuși așa-numitul „indice al deplasărilor structurale"

$$I_{z(q_l)} : \quad I_{z(q_{pl})} = \frac{0{,}85}{0{,}85} = 1$$

ne arată lipsa oricărei influențe a deplasării în asortimentul producției asupra schimbării prețului de cost. Este evident că în exemplul ales,

$$\rho_{i_z i_q} = 0$$

și că numai cu ajutorul acestui coeficient putem interpreta corect rezultatul obținut prin calculul celor doi indici.

4. Corelația dintre indici ai mărimilor medii (de tip C*) $\left(I_{\alpha} = \frac{\bar{\alpha}_1}{\bar{\alpha}_0}\right)$ și indici de alt tip

Corelația dintre indicii de tip C și de tip B. Să presupunem că ne este dat indicele unei mărimi medii și indicele agregat construit cu ajutorul ponderilor din perioada curentă (să zicem că este vorba de indici ai prețurilor). Făcînd raportul dintre acești indici și apoi transformările cuvenite, obținem

$$I_{\bar{p}} : \frac{\Sigma p_1 q_1}{\Sigma p_0 q_1} = \frac{\frac{\Sigma p_1 q_1}{\Sigma q_1}}{\frac{\Sigma p_0 q_0}{\Sigma q_0}} : \frac{\Sigma p_1 q_1}{\Sigma p_0 q_1} = \frac{\Sigma p_0 q_1}{\Sigma p_0 q_0} : \frac{\Sigma q_1}{\Sigma q_0} = \frac{\Sigma \frac{q_1}{q_0} p_0 q_0}{\Sigma p_0 q_0} : \frac{\Sigma \frac{q_1}{q_0} q_0}{\Sigma q_0}.$$

Pentru a transforma, mai departe, partea dreaptă a egalității de mai sus, aplicînd formula (XVII, 2.1 a), vom considera

$$x = \frac{q_1}{q_0}, \quad M = p_0 q_0, \quad N = q_0$$

și în acest caz putem scrie

$$\frac{\Sigma \frac{q_1}{q_0} p_0 q_0}{\Sigma p_0 q_0} : \frac{\Sigma \frac{q_1}{q_0} q_0}{\Sigma q_0} = 1 + \rho_{p_0 ; \frac{q_1}{q_0}} \cdot V_{p_0} V_{\frac{q_1}{q_0}}. \qquad \text{(XVII, 4.1)}$$

*) Vezi clasificația indicilor de la pagina 366 și următoarele.

Rezultă clar că o condiție suficientă ca să avem îndeplinită egalitatea

$$I_{\bar{p}} = I_{p(q_1)}$$

este lipsa corelației dintre indicii individuali ai cantităților $i_q = \frac{q_1}{q_0}$ și mărimile caracteristicii p din *perioada de bază* (p_0); mai există apoi și alte două condiții suficiente: egalitatea tuturor indicilor individuali ai cantităților și egalitatea tuturor mărimilor p din perioada de bază (lipsa totală a variației acestui indicator).

Prin urmare, dacă

$$\rho_{p_0 i_q} \neq 0,\ V_{p_0} \neq 0 \text{ și } V_{\frac{q_1}{q_0}} \neq 0$$

avem neegalitatea

$$I_{\bar{p}} \neq I_{p(q_1)}.$$

Mărimea diferenței dintre acești indici este determinată de nivelul valorilor pe care le iau cele trei elemente componente:

1°. coeficientul de variație a caracteristicii p în perioada de bază;

2°. coeficientul de variație a indicilor individuali ai cantităților diferitelor produse;

3°. coeficientul de corelație dintre primele două variabile.

Totodată, dacă avem inegalitatea

$$\rho_{p_0 \frac{q_1}{q_0}} > 0,$$

ceea ce înseamnă că indicii individuali ai cantităților sînt direct proporționali*) cu valorile caracteristicii variabile din perioada de bază, avem

$$I_{\bar{p}} > I_{p(q_1)}$$

și invers, dacă se realizează inegalitatea

$$\rho_{p_0 i_q} < 0,$$

avem

$$I_{\bar{p}} < I_{p(q_1)}.$$

În mod analog se obțin și formele exprimînd raportul dintre indicele unei mărimi medii și indicele agregat, ponderat cu cantitățile din perioada de bază:

$$I_{\bar{p}} : I_{p(q_0)} = \frac{\frac{\Sigma p_1 q_1}{\Sigma q_1}}{\frac{\Sigma p_0 q_0}{\Sigma q_0}} : \frac{\Sigma p_1 q_0}{\Sigma p_0 q_0} = 1 + \rho_{p_1 \frac{q_1}{q_0}} V_{p_1} V_{\frac{q_1}{q_0}}. \qquad \text{(XVII, 4.1a)}$$

*) Aici termenii: „direct proporționali" și „invers proporționali" sînt folosiți în sensul corelației directe și corelației inverse și nu de dependență funcțională.

În acest caz, caracterul corelației dintre indicele mediu și indicele agregat este determinat de semnul coeficientului de corelație $\rho_{p_1 \frac{q_1}{q_0}}$.

Formulele (XVII, 4.1) și (XVII, 4.1a) au multiple aplicații, dat fiind că în practică adeseori se calculează indicii cu structură variabilă de tipul $I_{\bar{\alpha}} = \frac{\bar{\alpha}_1}{\bar{\alpha}_0}$ ca și indicii agregați cu structură fixă.

Raționînd în mod analog ca la determinarea formulelor (XVII, 4.1) și (XVII, 4.1 a) se pot stabili și alte formule exprimînd corelația dintre indicele mediei și orice alt indice de forma mediei aritmetice.

De exemplu, se poate determina corelația dintre indicele productivității muncii calculat după metoda producției globale cu indicele productivității muncii propus de acad. S. Strumilin.

Ambii indici ne sînt cunoscuți din capitolul XV.

Făcînd și în acest caz raportul respectiv și apoi toate transformările cuvenite, obținem

$$\frac{\frac{\Sigma Q_1}{\Sigma T_1} : \frac{\Sigma Q_0}{\Sigma T_0}}{\frac{\Sigma \frac{Q_1}{Q_0} T_0}{\Sigma T_1}} = \frac{\Sigma Q_1}{\Sigma Q_0} : \frac{\Sigma \frac{Q_1}{Q_0} T_0}{\Sigma T_0} = \frac{\Sigma \frac{Q_1}{Q_0} Q_0}{\Sigma Q_0} : \frac{\Sigma \frac{Q_1}{Q_0} T_0}{\Sigma T_0} = 1 + \rho_{\frac{Q_0}{T_0}\frac{Q_1}{Q_0}} V_{\frac{Q_0}{T_0}} V_{\frac{Q_1}{Q_0}}. \qquad \text{(XVII, 4.2)}$$

După cum se vede, una din următoarele trei condiții sînt suficiente pentru ca să avem egalitatea celor doi indici:

1) Același ritm de creștere al volumului fizic al producției la toate întreprinderile;

2) Același nivel al productivității muncii în perioada de bază (producția globală pe un muncitor), la toate întreprinderile.

3) Inexistența corelației liniare dintre ritmul de creștere a producției la diferite întreprinderi și nivelurile din perioada de bază ale productivității muncii la aceste întreprinderi.

Dacă avem inegalitatea

$$\rho_{\frac{Q_0}{T_0}\frac{Q_1}{T_1}} > 0,$$

atunci indicele productivității muncii calculat după metoda producției globale este mai mare decît celălalt indice propus de acad. S. Strumilin (în acest caz ritmul de creștere al producției la întreprinderile cu un înalt nivel al productivității muncii în perioada de bază este mai mare decît la întreprinderile cu un nivel scăzut al productivității muncii).

Invers, dacă avem ingalitatea

$$\rho_{\frac{Q_0}{T_0}\frac{Q_1}{Q_0}} < 0,$$

atunci indicele calculat după metoda acad. S. Strumilin este mai mare (în acest caz ritmul de creștere a producției este în relație inversă față de nivelul productivității muncii din perioada de bază).

Să examinăm exemplul din tabelul 81 dat de L. Kazineț.

Tabelul 81

Întreprinderi	Producția globală în prețuri constante		Numărul muncitorilor		Producția globală pe un muncitor		Indici:	
	Perioada de bază	Perioada curentă	Perioada de bază	Perioada curentă	Perioada de bază	Perioada curentă	ai volumului fizic al producției	ai numărului de muncitori
	1	2	3	4	5	6	7	8
A	200	220	60	54	3,33	4,08	1,1	0,9
B	500	600	30	39	16,67	15,38	1,2	1,3
C	300	480	50	60	6,00	8,00	1,6	1,2
Total	1 000	1 300	140	153	7,14	8,49	1,3	1,1

Avem:

a) indicele productivității muncii calculat după metoda producției globale:

$$I_{pr.m.} = \frac{1\,300}{153} : \frac{1\,000}{140} = \frac{182}{153} = 1{,}190 \text{ sau } 119\%;$$

b) indicele productivității muncii calculat după metoda acad. S. Strumilin:

$$I_{pr.m.} = \frac{\frac{1{,}1}{0{,}9}\cdot 54 + \frac{1{,}2}{1{,}3}\cdot 39 + \frac{1{,}6}{1{,}2}\cdot 60}{54 + 39 + 60} = \frac{182}{153} = 1{,}190 \text{ sau } 119\%.$$

Este evident că identitatea celor doi indici din acest exemplu se datorează inexistenței corelației liniare dintre ritmul creșterii producției și nivelul productivității muncii în perioada de bază; condiția a treia este îndeplinită:

$$\rho_{\frac{V_0}{T_0}\frac{V_1}{V_0}} = 0.$$

În sfîrșit, se mai poate prezenta încă un caz de aplicare a formulei (XVII, 4.1) (tabelul 82).

Tabelul 82

Întreprinderi	Prețul de cost unitar (*z*)		Producția în mii unități	
	Perioada de bază	Perioada curentă	Perioada de bază	Perioada curentă
1	10	9,5	13	15
2	5	4,5	14	2
3	4	2,8	5	15

Se observă că de la o perioadă la alta au avut loc schimbări bruște în structura colectivității de produse ca și în nivelul prețului de cost. Cu toate acestea, indicele medici prețului de cost este egal cu indicele agregat cu structură fixă (ponderile perioadei curente), ceea ce înseamnă că indicele deplasărilor structurale este egal cu 1.

Într-adevăr, avem

$$I_{\bar{z}} = \frac{\bar{z}_1}{\bar{z}_0} = \frac{(9{,}5\cdot 15) + (4{,}5\cdot 2) + (2{,}8\cdot 15)}{32} : \frac{(10\cdot 13) + (5\cdot 14) + (4\cdot 5)}{32} =$$

$$= \frac{193}{220} = 0{,}88 \text{ sau } 88\%$$

și

$$I_{z(q_1)} = \frac{\Sigma z_1 q_1}{\Sigma z_0 q_1} = \frac{(9{,}5\cdot 15) + (4{,}5\cdot 2) + (2{,}8\cdot 15)}{(10\cdot 15) + (5\cdot 2) + (4\cdot 15)} = \frac{193{,}5}{220} = 0{,}88 \text{ sau } 88\%.$$

Prin urmare, avînd

$$I_{\text{str.}} = I_{\bar{z}} : I_{z(q_1)} = 1 \text{ sau } 100\%,$$

putem conchide că deplasările structurale n-au avut nici o influență asupra schimbării prețului de cost mediu. Aceste fapt se explică prin inexistența corelației liniare dintre indicii individuali ai volumului fizic al diferitelor produse și nivelurile corespunzătoare ale prețului de cost al acestor produse în perioada de bază.

Aceasta este o condiție suficientă ca indicele deplasării structurale să fie egal cu 1.

O contribuție prețioasă în studiul raportului dintre diferiți indici o aduce L. Tövissi *), abordînd, printre altele, corelația dintre doi indici ai productivității muncii cu structură fixă. Vom expune pe scurt considerațiile acestuia.

Este examinat raportul :

$$\frac{\Sigma w_1 T_1}{\Sigma w_0 T_1} : \frac{\Sigma q_1 t_0}{\Sigma q_1 t_1}$$

care, după o transformare simplă, se poate scrie

$$\frac{\Sigma i_w w_0 T_1}{\Sigma w_0 T_1} : \frac{\Sigma i_w T_1}{\Sigma T_1}.$$

Aplicînd formula (XVII, 2.1a) se obține

$$\frac{\Sigma i_w w_0 T_1}{\Sigma w_0 T_1} : \frac{\Sigma i_w T_1}{\Sigma T_1} = 1 + \rho_{i_w w_0} \cdot V_{i_w} \cdot V_{w_0}.$$

*) L. Tövissi, *Aplicarea coeficientului de corelație în teoria indicilor*. Revista de Statistică, 4 (1961).

După cum se vede, cei doi indici ai productivității muncii cu structură fixă vor fi egali între ei atunci cînd una din următoarele condiții este îndeplinită:

1) coeficientul de corelație $\rho_{i_w w_0}$ dintre indicii individuali ai productivității muncii și nivelurile individuale ale productivității muncii în perioada de bază este egal cu zero;

2) coeficientul de variație V_{i_w} a indicilor individuali ai productivității muncii este egal cu zero (ceea ce presupune că i_w — constant, adică productivitatea muncii la fiecare compartiment de muncă crește în același ritm);

3) coeficientul de variație V_{w_0} a nivelurilor productivității muncii în perioada de bază este egal cu zero (pentru care este necesar ca w_0 = constant, adică este necesar ca nivelul productivității muncii în perioada de bază să fie același la fiecare compartiment de muncă).

În general, pentru ca cei doi indici ai productivității muncii cu structură fixă să fie egali, este necesar să nu existe corelație liniară între creșterea productivității muncii (caracterizată prin indicii individuali) și nivelurile individuale ale productivității muncii în perioada de bază. În caz contrar, cei doi indici se deosebesc între ei.

Dacă coeficientul de corelație dintre indicii individuali ai productivității muncii și nivelurile individuale ale productivității muncii în perioada de bază este pozitiv, adică dacă productivitatea muncii crește mai rapid acolo unde nivelul ei de bază a fost mai ridicat, atunci indicele productivității muncii cu structură fixă, calculat pe baza indicatorului direct (cantitatea de produse în unitatea de timp de muncă w) va fi mai mare decît cel calculat pe baza indicatorului invers (consumul specific de timp de muncă pe unitatea de produs, t) și, dimpotrivă, dacă coeficientul de corelație este negativ, adică dacă productivitatea muncii crește mai rapid acolo unde nivelul ei în perioada de bază a fost mai scăzut, atunci al doilea indice va fi mai mare decît primul.

Autorul examinează și exemplul din tabelul 83.

Tabelul 83

Raioane	Perioada de bază			Perioada curentă			Indicii individuali ai productivității muncii i_w
	Extracția de cărbune (tone) q_0	Numărul muncitorilor T_0	Producția medie pe muncitor w_0	Extracția de cărbune (tone) q_1	Numărul muncitorilor T_1	Producția medie pe muncitor w_1	
I	6 400 000	8 000	800	10 800 000	12 000	900	1,125
II	7 200 000	12 000	600	5 600 000	8 000	700	1,167
Total	13 600 000	20 000	680	16 400 000	20 000	820	1,2058

Scanned with OKEN Scanner

Indicele cu structură fixă al productivității muncii pe baza metodei directe este

$$\frac{\Sigma w_1 T_1}{\Sigma w_0 T_1} = \frac{16\,400\,000}{12\,000 \times 800 + 8\,000 \times 600} =$$

$$= \frac{16\,400\,000}{14\,400\,000} = 1{,}13889 \text{ sau în procente } 113{,}98\%,$$

iar indicele cu structură fixă pe baza metodei indirecte:

$$\frac{\Sigma q_1 t_0}{\Sigma q_1 t_1} = \frac{10\,800\,000 \dfrac{8\,000}{6\,400\,000} + 5\,600\,000 \dfrac{12\,000}{7\,200\,000}}{20\,000} =$$

$$= \frac{13\,500 + 9\,333{,}33}{20\,000} = \frac{22\,833{,}33}{20\,000} = 1{,}14167 \text{ sau în procente, } 114{,}17\%.$$

Cei doi indici, într-adevăr, nu sînt egali. Raportul lor se poate explica cu ajutorul relației

$$\frac{\Sigma w_1 T_1}{\Sigma w_0 T_1} : \frac{\Sigma q_1 t_0}{\Sigma q_1 t_1} = \frac{1{,}13889}{1{,}14167} = 0{,}99757 = 1 + \rho_{i_w, w_0} \cdot V_{i_w} \cdot V_{w_0}.$$

Calculul coeficientului de corelație și al celor doi coeficienți de variație se poate efectua pe baza formulelor cunoscute, după unele transformări elementare:

$$\rho_{i_w w_0} = \frac{\Sigma q_1 \Sigma T_1 - \Sigma w_0 T_1 \Sigma q_1 t_0}{\sqrt{\Sigma T_1 \Sigma w_0^2 T_1 - (\Sigma w_0 T_1)^2]\,[\Sigma T_1 \Sigma i_w^2 T_1 - (\Sigma q_1 t_0)^2]}}$$

$$V_{w_0} = \frac{\sqrt{\Sigma T_1 \Sigma w_0^2 T_1 - (\Sigma w_0 T_1)^2}}{\Sigma w_0 T_1}$$

$$V_{i_w} = \frac{\sqrt{\Sigma T_1 \Sigma i_w^2 T_1 - (\Sigma q_1 t_0)^2}}{\Sigma q_1 t_0}.$$

Pentru a determina cantitățile care intervin în formule se alcătuiește tabelul 84.

Tabelul 84

Raioane	$w_0 T_1$	$q_1 t_0$	$w_0^2 T_1$	i_w^2	$i_w^2 T_1$
I	9 600 000	13 500	7 680 000 000	1,265625	15 187,5
II	4 800 000	9 333,33	2 880 000 000	1,361111	10 888,89
Total	14 400 000	22 833,33	10 560 000 000	2,626736	26 076,39

Făcînd înlocuirile în formulele de mai sus, se obține;

$$\rho_{i_w w_0} =$$

$$= \frac{16\,400\,000 \times 20\,000 - 14\,400\,000 \times 22\,833{,}33}{\sqrt{[20\,000 \times 10\,560\,000\,000 - 14\,400\,000^2][20\,000 \times 26\,076{,}39 - 22\,833{,}33^2]}} =$$

$$= -0{,}99941$$

$$V_{w_0} = \frac{\sqrt{20\,000 \times 10\,560\,000\,000 - 14\,400\,000^2}}{14\,400\,000} = 0{,}13608,$$

$$V_{i_w} = \frac{\sqrt{20\,000 \times 26\,076{,}39 - 22\,833{,}33^2}}{22\,833{,}33} = 0{,}017889.$$

Folosind aceste calcule, avem

$$\frac{1{,}13889}{1{,}14167} = 1 - 0{,}99941 . 0{,}13608 . 0{,}017889 = 1 - 0{,}002433 = 0{,}99757.$$

Aşadar, diferența dintre cei doi indici în exemplul analizat se pune pe seama celor doi coeficienți de variație $V_{w_0} = 13{,}608\%$ şi $V_{i_w} = 1{,}789\%$, întrucît coeficientul de corelație în valoare absolută este aproximativ egal cu 1. Indicele calculat pe baza indicatorului direct este mai mic decît indicele calculat pe baza indicatorului invers, deoarece între indicii individuali ai productivității muncii şi nivelul productivității muncii în perioada de bază există o corelație liniară negativă strînsă. Diferența indicilor, deşi există legătura strînsă, nu este mare, deoarece variația indicilor individuali ai productivității muncii şi a nivelurilor ei individuale este mică ($V_{w_0} = 13{,}6\%$, respectiv $V_{i_w} = 1{,}8\%$), ceea ce face ca diferența să fie practic neînsemnată.

Scanned with OKEN Scanner

Scanned with OKEN Scanner

Capitolul XVIII

Aplicarea legii numerelor mari la elaborarea indicilor

1. Considerații introductive

La analiza statistico-economică a fenomenelor social-economice, metoda indicilor capătă o utilizare tot mai mare. Numeroase probleme — în special cele legate de relevarea influenței diferiților factori asupra fenomenului complex — se pun cu prilejul acestei analize. Ele se complică și devin deosebit de subtile în cazul construirii indicilor medii.

De pildă, o problemă importantă care se pune este aceea dacă la efectuarea analizelor prin metoda indicilor trebuie să ținem seama și de *variația întîmplătoare* a indicatorilor, ce se manifestă, de obicei, în cazul cînd este supus observării un număr mic de unități, știut fiind că o asemenea variație nu este neglijată ori de cîte ori în cadrul analizelor se aplică valorile medii sau ceilalți indicatori sintetici (ai dispersiei, corelației etc.).

De această problemă s-a ocupat în mod special I. I. Lukomski *), luînd drept obiect de studiu indicele productivității muncii sub forma mediei aritmetice a indicilor individuali, ponderați cu cheltuielile de timp de muncă din perioada curentă. Acest autor observă că, întrucît și indicii sînt indicatori generalizatori, ei pot fi folosiți ca mijloc de cercetare a legității statistice. Se știe, însă, că o legitate statistică se manifestă evident numai în cazul cînd este supus observării un mare număr de unități de

*) I. I. Lukomski (1906—1956), un remarcabil specialist în statistica matematică, doctor în științe economice. În activitatea sa științifică și didactică, I. I. Lukomski s-a ocupat îndeosebi de problemele legate de aplicațiile practice ale statisticii matematice în producție.

Problema aplicării legii numerelor mari la elaborarea indicilor a fost abordată de I. I. Lukomski într-un studiu intitulat: *Teoria korreliații i ee primenenie k analizu proizvodstva*. Gosstatizdat, Moskva, 1958.

același tip și că în cazul observării unui număr mic de unități ea este estompată prin influența factorilor întîmplători.

Dacă, de pildă, vrem să ilustrăm ce anume legitate statistică se manifestă cînd comparăm dinamica productivității muncii la diferite întreprinderi cu grad diferit de mecanizare a muncii sau de organizarea producției etc., este necesar să considerăm un număr destul de mare de muncitori în ambele perioade.

Indicii productivității muncii — la fel ca și valorile medii sau alți indicatori sintetici — calculați pe baza observării unui mic număr de muncitori, au un caracter întîmplător, nereprezentativ pentru legitatea statistică ce ne interesează.

Erorile întîmplătoare ale indicilor se supun teoriei legii numerelor mari și de aceea, în lumina ei, trebuie examinate unele probleme ale construirii indicilor.

Pentru ilustrarea acestor teze I. I. Lukomski consideră următorul exemplu (vezi tabelele 85 și 86).

Tabelul 85

Complexul I (grupă de mașini-unelte)				
Sectoare	Producția pe un muncitor (în unități naturale)		Numărul muncitorilor	
	Perioada de bază	Perioada curentă	Perioada de bază	Perioada curentă
1	76	57	1	100
2	50	62	100	100

Notă. În cadrul acestui complex în ambele sectoare s-au introdus o serie de măsuri de raționalizare în perioada curentă.

Tabelul 86

Complexul II (grupă de mașini-unelte)				
Sectoare	Producția pe un muncitor (în unități naturale)		Numărul muncitorilor	
	Perioada de bază	Perioada curentă	Perioada de bază	Perioada curentă
1	50	64	1	99
2	70	72	50	50

Notă. În cadrul complexului II nu s-a luat nici o măsură de raționalizare.

După cum se vede din tabele, datele concrete au fost potrivite în așa fel ca ideea autorului să fie scoasă în evidență cît mai pregnant.

Scanned with OKEN Scanner

1) Calculul indicilor individuali ai productivității muncii, pe sectoare, ne dă următoarele rezultate:

Pentru complexul I:

$$\text{Sectorul 1} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad I_1 = \frac{57}{76} = 0{,}75,$$

$$\text{Sectorul 2} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad I_2 = \frac{62}{50} = 1{,}24.$$

Pentru complexul II:

$$\text{Sectorul 1} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad I_1 = \frac{64}{50} = 1{,}28,$$

$$\text{Sectorul 2} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad I_2 = \frac{72}{70} = 1{,}03.$$

2) Calculul indicilor medii ai productivității muncii, luînd drept ponderi numărul muncitorilor din perioada curentă (acest procedeu de ponderare este în general admis, sarcina economică fiind aceea de a releva economia de timp de muncă realizată în perioada curentă ca rezultat al schimbării nivelurilor individuale ale productivității muncii față de perioada de bază), pentru fiecare complex în parte ne dă mărimile:

Pentru complexul I:

$$\bar{I}_{\text{I}} = \frac{0{,}75 \cdot 100 + 1{,}24 \cdot 100}{200} = 0{,}995.$$

Pentru complexul II:

$$\bar{I}_{\text{II}} = \frac{1{,}28 \cdot 99 + 1{,}03 \cdot 50}{149} = 1{,}20.$$

Se observă că, la complexul I, productivitatea medie *a scăzut* cu 0,5% (cu toate că în cadrul acestui complex s-au introdus măsuri de raționalizare a muncii), iar la complexul II, *a crescut* cu 20%.

Cel ce ar analiza aceste rezultate ar fi evident nedumerit, fiind înclinat a conchide că măsurile de raționalizare luate în perioada curentă în cadrul complexului I nu numai că n-au avut nici un fel de eficacitate economică, ci dimpotrivă au produs chiar un efect dăunător; aceasta s-ar pune în evidență cu atît mai mult, dacă din datele analizei economice ar reieși că ceilalți factori care influențează nivelul productivității muncii nu s-au schimbat sau indică o tendință firească de dezvoltare progresivă.

Or, este evident că o asemenea concluzie nu este fundamentată din punct de vedere statistic; ea este sugerată greșit numai datorită faptului că valorile individuale au fost greșit introduse în formula indicelui mediu al productivității muncii.

Într-adevăr — arată I. I. Lukomski — este uşor de observat că indicii individuali *) ai productivităţii muncii nu pot avea aceeaşi valoare din punct de vedere statistic şi de aceea nu li se pot atribui aceleaşi ponderi: indicatorii nivelului productivităţii muncii din cadrul unor sectoare în perioada de bază au fost obţinuţi numai de la un singur muncitor; în consecinţă, mărimea lor (76 unităţi în cadrul complexului I şi 50 unităţi în cadrul complexului II) a fost cu totul întîmplătoare şi nesemnificativă pentru a releva legitatea ce se manifestă în activitatea întreprinderii.

Rezultă de aici că şi indicele mediu, de pildă, cel calculat pentru complexul I (0,995), este nereprezentativ, întrucît el oglindeşte raportul unei mărimi reprezentative din perioada curentă (în care s-au considerat 100 muncitori), faţă de o mărime întîmplătoare, nereprezentativă, din perioada de bază.

Concluzia pe care o trage Lukomski este aceea că ponderea ce se impune a fi luată în considerare la construirea indicelui mediu al productivităţii muncii este alta, şi anume *o pondere determinată de numărul muncitorilor din ambele perioade*. Pe de altă parte, autorul observă că nu poate fi ignorat nici acel sistem de ponderi care se impune în mod firesc din condiţiile concrete ale problemei economice supuse studiului ca şi din funcţia pe care indicele construit este chemat a o îndeplini.

Problema constă, deci, în a găsi un sistem de ponderi care să combine just premizele economice ce stau la baza unui indice, cu acele consideraţii ce decurg din cerinţa reprezentativităţii indicilor (în sensul micşorării influenţei erorilor întîmplătoare).

Problema pusă de I. I. Lukomski este foarte importantă. Considerarea ponderii ca o funcţie de numărul muncitorilor din ambele perioade conduce la o ameliorare a sistemului de indici existenţi. Vom expune pe scurt metoda folosită de el pentru rezolvarea problemei, cu scopul de a o supune apoi la o analiză mai amplă şi a trage unele concluzii pe marginea acestei importante probleme.

2. Sistemul reprezentativ

În ipoteza în care ne propunem să eliminăm influenţa erorilor întîmplătoare, consideraţiile cu caracter economic care stau la baza alegerii ponderilor trebuie să le legăm nemijlocit de valorile *teoretice* ale indicilor.

Prin valori teoretice, adică deduse teoretic, I. I. Lukomski înţelege acele valori ale indicilor care exprimă exact legitatea procesului examinat; cu alte cuvinte, este vorba de valori ale indicilor care sînt eliberate de erori întîmplătoare (de influenţa factorilor întîmplători) şi se pot manifesta numai în cazul observării unui număr cît mai mare de unităţi statistice (muncitori). Sistemul de ponderi corespunzător acestor valori teoretice

*) Prin indicii individuali se înţeleg aici nu numai indicii de forma $\frac{\alpha_1}{\alpha_0}$ ci şi indicii de forma $\frac{\bar{\alpha}_1}{\bar{\alpha}_0}$.

este denumit *sistem de fixare* și se notează cu $\alpha_1, \alpha_2, \ldots, \alpha_s$. Pe baza acestui sistem de ponderi și a valorilor teoretice ale indicilor individuali se calculează *valoarea teoretică a indicelui mediu.*

Cît privește mărimile efective ale indicilor cu care operează statistica, ele trebuie ponderate cu ajutorul unui alt sistem de ponderi $g_1, g_2, \ldots, g_s$, care trebuie construit tot pe baza sistemului de fixare a ponderilor ținîndu-se seama însă de cerința reprezentativității (în sensul legii numerelor mari).

I. I. Lukomski formulează în felul următor principiul reprezentativității:

Valoarea efectivă a indicelui mediu, obținută din valorile efective ale indicilor individuali, cu ajutorul ponderilor g_i, trebuie să fie foarte apropiată de valoarea teoretică a indicelui mediu obținută din valorile teoretice ale indicilor individuali cu ajutorul ponderilor α_i.

Lukomski consideră că în acest principiu își găsește reazim calea justă de a combina premisele economice ale formulei indicelui cu cerințele reprezentativității acestuia. Pe baza lui se deduc și ponderile $g_1, g_2, \ldots, g_s$, care constituie tocmai *sistemul reprezentativ de ponderi.*

Prin urmare, problema ponderării indicelui mediu se subdivide în două:

1) Găsirea sistemului de fixare al ponderilor ($\alpha_1, \alpha_2, \ldots, \alpha_s$) cu ajutorul căruia se calculează apoi valoarea medie teoretică a indicelui, din valorile teoretice ale indicilor individuali *).

2) Găsirea sistemului de ponderi ($g_1, g_2, \ldots, g_s$). Tocmai de această problemă se ocupă I. I. Lukomski care o rezolvă pe baza principiului reprezentativității formulat mai sus.

La rezolvarea acestei probleme se ivesc două cazuri distinctive:

A. Variația indicilor individuali efectivi este mică, fapt care justifică ca la baza acestor indici să fie pusă una și aceeași valoare teoretică.

B. Variația indicilor individuali efectivi este destul de mare ceea ce necesită ca la baza lor să fie puse valori teoretice diferite. Acesta este, de altfel, cazul general.

Vom arăta în continuare procedeul de calcul al ponderilor g_i într-un caz și în altul ca și criteriile practice pentru identificarea acestora, elaborate de I. I. Lukomski.

Se folosește următoarea notație:

$n_1, n_2, \ldots, n_s$, numărul muncitorilor în perioada de bază, în sectoarele I,II etc., (vezi tabelele 85 și 86) ale complexului de producție;

$n_1', n_2', \ldots, n_s'$, numărul muncitorilor în perioada curentă (idem);

$\bar{x}_1 = \frac{\Sigma_1 x}{n_1}$; $\bar{x}_2 = \frac{\Sigma_2 x}{n_2}, \ldots$, producția medie a unui muncitor în perioada de bază, în sectoarele respective ale complexului;

$\bar{x}_1' = \frac{\Sigma_1 x'}{n_1'}$; $\bar{x}_2' = \frac{\Sigma_2 x'}{n_2'}, \ldots$, producția medie a unui muncitor în perioada curentă,

*) Această problemă avînd o importanță generală pentru teoria indicilor nu este examinată aici.

în sectoarele respective ale complexului;

$I_1 = \frac{\bar{x}_1'}{\bar{x}_1}$; $I_2 = \frac{\bar{x}_2'}{\bar{x}_2}, \ldots$, indicii individuali ai productivității muncii în diferite sectoare ale complexului;

$\bar{I} = \frac{g_1 I_1 + g_2 I_2 + \ldots + g_s I_s}{g_1 + g_2 + \ldots + g_s}$, indicele mediu al productivității muncii, în care $g_1, g_2, \ldots, g_s$ sînt ponderile căutate.

A. *Cazul cînd la baza indicilor individuali avem aceeași valoare teoretică* (I_0). Să admitem că la baza indicilor individuali $I_1, I_2, \ldots, I_s$ avem una și aceeași valoare teoretică I_0. În acest caz ponderile sistemului de fixare ($\alpha_1, \alpha_2, \ldots, \alpha_s$) ne sînt indiferente, dat fiind că orice sistem al ponderilor (α_i) ne duce la aceeași valoare teoretică I_0 a indicelui mediu $\bar{I}$, fiind totodată și valoarea teoretică a fiecărui indice I_i. Singurul scop al sistemului reprezentativ de ponderi ($g_1, g_2, \ldots, g_s$), în acest caz, este acela de a realiza o apropiere cît mai mare a indicelui mediu $\bar{I}$ de valoarea teoretică I_0.

Să notăm cu $\mathrm{d}\bar{I}$ eroarea întîmplătoare a indicelui $\bar{I}$, adică abaterea lui $\bar{I}$ de la speranța sa matematică $E\bar{I}$ (indicele teoretic mediu $\bar{I}_0$)*)

Reprezentativitatea indicelui $\bar{I}$ o vom măsura cu dispersia lui stohastică

$$\sigma_{\bar{I}}^2 = E(\mathrm{d}\bar{I})^2.$$

Cerința principiului reprezentativității poate fi formulată matematic în felul următor:

Să se afle un asemenea sistem de ponderi $g_1, g_2, \ldots, g_s$ *pentru indicele mediu* $\bar{I} = \frac{\Sigma g_i I_i}{\Sigma g_i}$, *care să satisfacă condiția ca dispersia* $\sigma_{\bar{I}}^2$ *să fie minimă.*

Această problemă I. I. Lukomski o rezolvă după cum urmează:

Fie $\gamma_j = \frac{g_j}{\Sigma g_j}$ „greutățile specifice"; atunci

$$\Sigma\gamma_i = 1, \quad \bar{I} = \Sigma\gamma_i I_i, \quad I_0 = \Sigma\gamma_i I_{i_0},$$

$$\sigma_{\bar{I}}^2 = E(\mathrm{d}\bar{I})^2 = E(\bar{I} - I_0)^2 = E[\Sigma\gamma_i(I_i - I_{i0})]^2 = \Sigma\gamma_i^2\sigma_{I_i}^2, \qquad \text{(XVIII, 2.1)}$$

(întrucît erorile întîmplătoare ale indicilor individuali $\mathrm{d}I_i = I_i - I_{i0}$ și $\mathrm{d}I_j = I_j - I_{j0}$ pot fi considerate independente).

Cerința minimului este:

$$F = \sigma_{\bar{I}}^2 = \Sigma\gamma_i^2\sigma_{I_i}^2 \text{ minim.} \qquad \text{(XVIII, 2.2)}$$

cu condiția $\Sigma\gamma_i = 1$.

*) Aici, I. I. Lukomski face abstracție de o oarecare deplasare a speranței matematice EI (sau $E\bar{I}$) față de valoarea teoretică (la limită) I_0. Această deplasare trebuie considerată, practic, neînsemnată, devenind neglijabilă odată cu creșterea numărului de cazuri observate, dacă mărimea de la numitorul indicelui este mult diferită de zero.

Scanned with OKEN Scanner

Condiţiile necesare şi suficiente pentru minimul condiţionat al formei pătratice F pozitiv definită sînt :

$$\frac{\partial F}{\partial \gamma_1} = \frac{\partial F}{\partial \gamma_2} = \ldots = \frac{\partial F}{\partial \gamma_s}, \qquad \Sigma \gamma_i = 1$$

sau

$$\gamma_1 \sigma_{I_1}^2 = \gamma_2 \sigma_{I_2}^2 = \ldots = \gamma_s \sigma_{I_s}^2, \qquad \Sigma \gamma_i = 1 \tag{XVIII, 2.3}$$

sau

$$g_1 \sigma_{I_1}^2 = g_2 \sigma_{I_2}^2 = \ldots = g_s \sigma_{I_s}^2, \tag{XVIII, 2.3'}$$

adică ponderile $g_1, g_2, \ldots, g_s$, în cazul indicilor individuali $I_1, I_2, \ldots, I_s$, trebuie să fie invers proporţionale dispersiilor stohastice ale acestora.

Deoarece coeficientul proporţionalităţii la determinarea ponderilor nu joacă nici un rol, putem considera că

$$g_i = \frac{1}{\sigma_{I_i}^2}, \qquad (i = 1, 2, \ldots, s). \tag{XVIII, 2.4}$$

Mai trebuie calculată valoarea dispersiei $\sigma_{I_i}^2$. Folosind în acest scop formula aproximativă de diferenţiere

$$\mathrm{d}I_i = \frac{x_{i_0} \mathrm{d}\bar{x}_i' - \bar{x}_{i_0}' \mathrm{d}\bar{x}_i}{(\bar{x}_{i_0})^2},$$

în care cu $\bar{x}_{i_0}$ şi $\bar{x}_{i_0}'$ notăm prescurtat speranţele matematice $E\bar{x}_i$ şi $E\bar{x}_i'$, obţinem

$$\sigma_{I_i}^2 = E(\mathrm{d}I_i)^2 = \frac{(\bar{x}_{i_0})^2 \sigma_{\bar{x}_i'}^2 + (\bar{x}_{i_0}')^2 \sigma_{\bar{x}_i}^2}{(\bar{x}_{i_0})^2}. \tag{XVIII, 2.5}$$

Am neglijat cu acest prilej dependenţa dintre $\mathrm{d}x_i$ şi $\mathrm{d}\bar{x}_i'$ (ceea ce este cu atît mai justificat cu cît perioadele considerate sînt mai îndepărtate şi cu cît mişcarea forţei de muncă este mai mare).

Mai departe, folosind formulele

$$\sigma_{\bar{x}_i}^2 = \frac{\sigma_{i_0}^2}{n_i} \quad \text{şi} \quad \sigma_{\bar{x}_i'}^2 = \frac{\sigma_{i_0'}^2}{n_i'}$$

în care $\sigma_{i_0}^2$ şi $\sigma_{i_0'}^2$ sînt valorile teoretice ale dispersiei privind producţia individuală a muncitorilor şi considerînd coeficienţii de variaţie

$$V_{i_0} = \frac{\sigma_{i_0}}{\bar{x}_{i_0}} \qquad \text{şi} \qquad V_{i_0}' = \frac{\sigma_{i_0}'}{\bar{x}_{i_0}'}$$

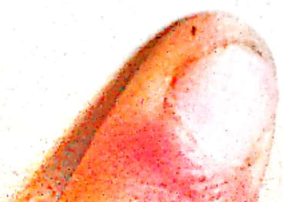
aproximativ egali, obţinem din formula (XVIII, 2.5)

$$\sigma_{I_i}^2 = \frac{2\left(x_{i_0}'\right)^2}{h_i} V_{i_0}^2 \tag{XVIII, 2.6}$$

în care

$$h_i = \frac{2}{\frac{1}{n_i} + \frac{1}{n_i'}} \tag{XVIII, 2.7}$$

este media armonică calculată pe baza numărului de muncitori din perioadele de bază şi curentă.



Scanned with OKEN Scanner

Substituind expresia (XVIII, 2.6) în formula (XVIII, 2.4), admițind că coeficienții de variație V_{i_0} diferă foarte puțin unul de altul și neglijînd coeficienții constanți la determinarea ponderilor, obținem definitiv

$$g_i = h_i\,, \qquad (i = 1, 2, \ldots, s) \qquad \text{(XVIII, 2.8)}$$

în care h_i este determinat de formula (XVIII, 2.7).

În ultimă analiză, I. I. Lukomski ajunge la următoarea formulă a indicelui mediu al productivității muncii:

$$\bar{I} = \frac{\Sigma h_i I_i}{\Sigma h_i}\,, \qquad \text{(XVIII, 2.9)}$$

în care h_i este media armonică dată de formula (XVIII, 2.7) servind drept ponderi pentru indicii individuali ai productivității muncii. După cum se vede, formula indicelui mediu al productivității muncii (XVIII, 2.9) este cea a mediei aritmetice ponderate. Această formulă — arată Lukomski — se bucură de reprezentativitate maximă, în sensul legii numerelor mari.

Aplicînd formula (XVIII, 2.9) la datele exemplelor din tabelele 85 și 86 *) se constată că indicele mediu al productivității muncii în sectoarele complexului I este egal cu 1,23, ceea ce înseamnă că s-a înregistrat o creștere a productivității muncii cu 23% și nu o scădere a ei cu 0,5%, așa cum ne-a arătat indicele mediu al productivității muncii calculat după metoda obișnuită. De asemenea, în sectoarele complexului II productivitatea muncii a crescut cu 4% și nu cu 20%.

Așadar, aplicînd formula (XVIII,2.9) am micșorat considerabil efectul denaturant al abaterii întîmplătoare a nivelurilor productivității muncii a unor muncitori izolați față de nivelul productivității muncii atins pe întreaga masă de muncitori. Obținînd în cadrul complexului I un indice al productivității muncii mai mare, sîntem îndreptățiți a trage concluzia ce se impunea, de altfel, în mod firesc, că măsurile de raționalizare luate în perioada curentă în cadrul complexului I au dat rezultate pozitive.

Verificarea preciziei indicelui productivității muncii propus de I. I. Lukomski (formula XVIII,2.9). Formula (XVIII,2.9) micșorează maximal influența denaturantă a factorilor întîmplători, însă nu o elimină complet întrucît asupra indicelui mediu mai exercită influență și erorile întîmplătoare ale indicilor individuali.

---

*) Media armonică pentru calculul indicelui s-a calculat astfel:

a) Pentru complexul I

$$h_1 = \frac{2}{\frac{1}{1} + \frac{1}{100}} \approx 2, \qquad h_2 = \frac{2}{\frac{1}{100} + \frac{1}{100}} = 100.$$

b) Pentru complexul II

$$h_1 = \frac{2}{\frac{1}{1} + \frac{1}{99}} \approx 2, \qquad h_2 = \frac{2}{\frac{1}{50} + \frac{1}{50}} = 50.$$

Eroarea medie a indicelui individual al productivității muncii se calculează după formula :

$$\sigma_{I_i} = V_0 I_0 \sqrt{\frac{2}{h_i}} \qquad \text{(XVIII, 2.6')}$$

ce se deduce din formula (XVIII, 2.6).

Eroarea medie a indicelui mediu $\bar{I}$ se calculează după formula

$$\sigma_{\bar{I}} = V_0 I_0 \sqrt{\frac{2}{\Sigma h_i}} \cdot \qquad \text{(XVIII, 2.10)}$$

Într-adevăr, am văzut că

$$\bar{I} = \frac{\Sigma h_i I_i}{\Sigma h_i} \cdot$$

Rezultă

$$\sigma^2_{\bar{I}} = \frac{\Sigma h_i^2 \sigma^2_{I_i}}{(\Sigma h_i)^2} \cdot$$

Însă din formula (XVIII, 2.6') rezultă

$$\sigma^2_{I_i} = I_0^2 V_0^2 \frac{2}{h_i} \cdot$$

Înlocuind valoarea lui $\sigma^2_{I_i}$ în relația de mai sus, rezultă valoarea lui $\sigma^{\bar{I}}$.

Potrivit datelor din exemplul considerat, aplicînd formula (XVIII,2.9) avem inegalitatea

$$\bar{I}_{\text{I}} > \bar{I}_{\text{II}} \qquad (\alpha)$$

Pentru a ne convinge că această inegalitate nu este întîmplătoare trebuie să demonstrăm că aceeași relație există și între valorile teoretice ale celor doi indici, adică

$$\bar{I}_{\text{I}_0} > \bar{I}_{\text{II}_0} \cdot \qquad (\beta)$$

Exactitatea acestei inegalități o putem verifica.

Să presupunem pentru moment că

$$\bar{I}_{\text{I}_0} = \bar{I}_{\text{II}_0}$$

și să demonstrăm că în această ipoteză inegalitatea (α) nu este probabilă. Prin aceasta, ipoteza

$$\bar{I}_{\text{I}_0} = \bar{I}_{\text{II}_0}$$

va fi infirmată și, implicit, fundamentată inegalitatea (β).

Diferența

$$d = \bar{I}_{\text{I}} - \bar{I}_{\text{II}} = 1{,}23 - 1{,}04 = 0{,}19$$

are o eroare medie care, admițînd ipoteza

$$I_{I_0} = I_{II_0} = I_0,$$

se calculează astfel:

$$\sigma_d = \sqrt{\sigma^2_{\bar{I}_I} + \sigma^2_{\bar{I}_{II}}} = \frac{2V_0 I_0}{\sqrt{H}}, \qquad \text{(XVIII, 2.11)}$$

în care

$$H = \frac{2}{\frac{1}{\Sigma_I h_i} + \frac{1}{\Sigma_{II} h_i}}.$$

Valoarea aproximativă a lui I_0 din formula (XVIII, 2.11) este următoarea:

$$I_0 \approx \frac{\bar{I}_I \Sigma_I h_i + \bar{I}_{II} \Sigma_{II} h_i}{\Sigma_I h_i + \Sigma_{II} h_i}. \qquad \text{(XVIII, 2.12)}$$

Să efectuăm calculele:

$$I_0 \approx \frac{1{,}23 \cdot 102 + 1{,}04 \cdot 52}{154} = 1{,}17,$$

$$\sqrt{H} = \sqrt{\frac{2}{\frac{1}{102} + \frac{1}{52}}} = \sqrt{69} = 8{,}3.$$

Pentru calculul erorii medii a lui d să presupunem că în exemplul considerat coeficientul de variație V_0 este aproximativ egal cu 0,20. Vom avea deci

$$\sigma_d = \frac{2 \cdot 0{,}20 \cdot 1{,}17}{8{,}3} = 0{,}057.$$

Notînd cu t raportul dintre d și σ_d, avem

$$t = \frac{d}{\sigma_d} = \frac{0{,}19}{0{,}057} = 3{,}33.$$

După cum se vede, diferența d este de peste trei ori mai mare decît eroarea sa medie teoretică. Aceasta înseamnă că practic, probabilitatea ca să avem $t > 3$ este egală cu zero. Prin urmare, diferența dintre $\bar{I}_I$ și $\bar{I}_{II}$ egală cu 0,19 nu s-a datorat unui joc al întîmplării și de aceea ipoteza care presupunea egalitatea $\bar{I}_{I0} = \bar{I}_{II0}$ trebuie respinsă.

Se demonstrează astfel că inegalitatea (β) este adevărată.

Determinarea criteriului de variabilitate. Am văzut că formula (XVIII, 2.9) a fost dedusă presupunînd că la baza tuturor indicilor individuali avem una și aceeași valoarea teoretică I_0 și că variabilitatea indicilor I_i este determinată de cazurile observate în număr mic. Este necesar, de

aceea, ca în lucrările practice să avem un criteriu sigur de a verifica că sîntem în cazul ipotezei făcute.

Folosind notația:

D_1, dispersia indicilor individuali:

$$D_1 = \frac{\Sigma h_i (\bar{I}_i - \bar{I})^2}{\Sigma h_i}: \qquad \text{(XVIII,2.13)}$$

D_{10}, speranța matematică a dispersiei D_I:

$$D_{10} = ED_I = \frac{2I_0^2 V_0^2 (s-1)}{\Sigma h_i}, \qquad \text{(XVIII,2.14)}$$

în care s este numărul indicilor individuali;

σ_{D_1}, eroarea medie pătratică a dispersiei D_I:

$$\sigma_{D_1} = \frac{2I_0^2 V_0^2}{\Sigma h_i} \sqrt{2(s-1)}, \qquad \text{(XVIII,2.15)}$$

putem considera că sîntem în ipoteza unei variații mici dacă avem inegalitatea

$$\frac{D_I}{D_{10}} < 1 + 3\frac{\sigma_{D_1}}{\sigma_{D_{10}}}$$

în care 3 reprezintă coeficientul de siguranță folosit de obicei în lucrările practice.

Criteriul de variabilitate K_1 propus de I. I. Lukomski este dat de relația

$$K_1 = \frac{\frac{D_I}{D_{10}} - 1}{\frac{\sigma_{DI}}{\sigma_{D_{10}}}} < 3$$

sau

$$K_1 = \left(\frac{D_1}{D_{10}} - 1\right)\sqrt{\frac{s-1}{2}} < 3. \qquad \text{(XVIII,2.16)}$$

În lucrările practice se poate calcula valoarea aproximativă a lui K_1, introducînd în formula (XVIII,2.14) în locul lui I_0, valoarea lui $\bar{I}$, iar în locul lui V_0, valoarea medie efectivă a coeficientului de variație:

$$K_1 = \left(\frac{\Sigma h_i (I_i - \bar{I})^2}{2I_0^2 V_0^2 (s-1)} - 1\right)\sqrt{\frac{s-1}{2}} < 3. \qquad \text{(XVIII, 2.16')}$$

Aplicarea formulei (XVIII, 2.9) a indicelui productivității muncii în cazul exemplului considerat este pe deplin justificată dat fiind că valoarea lui K_1 în complexele I și II este mai mică de 3.

Într-adevăr :
Pentru complexul I :

$$I_1 = 0{,}75, \qquad I_2 = 1{,}24,$$
$$h_1 = 2, \qquad h_2 = 100,$$
$$\bar{I}_I = 1{,}23,$$

V_0 să presupunem că este egal cu 0,2

$$\frac{D_1}{D_{10}} = \frac{\Sigma h_i (I_i - \bar{I})^2}{2 I_0^2 V_0^2 (s-1)},$$

$$\frac{D_1}{D_{10}} \approx \frac{2(0{,}75 - 1{,}23)^2 + 100(1{,}24 - 1{,}23)^2}{2 \cdot 1{,}23^2 \cdot 0{,}2^2} = \frac{0{,}47}{0{,}12} = 3{,}9,$$

$$K_1 = (3{,}9 - 1)\sqrt{0{,}5} \approx 2{,}0 < 3.$$

Pentru complexul II :

$$I_1 = 1{,}28, \qquad I_2 = 1{,}03,$$
$$h_1 = 2, \qquad h_2 = 50,$$
$$\bar{I}_{II} = 1{,}04,$$

$$V_0 \approx 0{,}2,$$

$$\frac{D_1}{D_{10}} = \frac{2(1{,}28 - 1{,}04)^2 + 50(1{,}03 - 1{,}04)^2}{2 \cdot 1{,}04^2 \cdot 0{,}2^2} = \frac{0{,}120}{0{,}086} = 1{,}4$$

$$K_1 = (1{,}4 - 1)\,0{,}7 = 0{,}28 < 3.$$

Prin urmare, ipoteza că la baza fiecărui indice individual efectiv avem una și aceeași valoare teoretică se verifică.

Dacă valoarea lui K_1 ar fi mai mare de 3, aceasta ar însemna că sîntem în cazul general, cînd la baza indicilor individuali efectivi stau valori teoretice diferite — cazul B.

B. *Cazul cînd la baza indicilor individuali avem valori teoretice diferite.* În acest caz, sistemului reprezentativ de ponderi $(g_1, g_2, \ldots, g_s)$ trebuie să i se aducă o anumită corecție.

Fie $\alpha_1, \alpha_2, \ldots, \alpha_s$ sistemul de fixare a ponderilor :

$$\Sigma \alpha_i = 1, \quad \alpha_j = \frac{\alpha_j}{\Sigma a_i}$$

și $\gamma_1, \gamma_2, \ldots, \gamma_s$ sistemul reprezentativ de ponderi corectate pe care-l căutăm :

$$\Sigma \gamma_i = 1, \quad \left(\gamma_j = \frac{g_i}{\Sigma g_i}\right)$$

$\bar{I}_0 = \Sigma \alpha_i I_{i_0}$, $\bar{I} = \Sigma \gamma_i I_i$ valorile — teoretică și efectivă — ale indicelui mediu.

Sistemul de ponderi căutat $\gamma_1, \gamma_2, \ldots, \gamma_s$ trebuie să satisfacă următoarea condiție de reprezentativitate:

$$F = E(\bar{I} - \bar{I}_0)^2 = \text{minim} \qquad \text{(XVIII, 2.17)}$$

sau, în altă formă,

$$F = E(\gamma_i I_i - \Sigma\alpha_i I_{i_0})^2 = E[\Sigma\gamma_i(I_i - I_{i_0}) + \Sigma I_{i_0}(\gamma_i - \alpha_i)]^2 = \\ = \Sigma\gamma_i^2\sigma_{I_i}^2 + [\Sigma I_{i_0}(\gamma_i - \alpha_i)]^2. \qquad \text{(XVIII, 2.18)}$$

Condițiile de minim ale funcției F ne duc la următorul sistem de ecuații cu necunoscutele $\gamma_1, \gamma_2, \ldots, \gamma_s$:

$$\left.\begin{aligned} &\gamma_j\sigma_{I_j}^2 + I_{j_0}\sum_{i=1}^{s} I_{i_0}(\gamma_i - \alpha_i) = \gamma_s\sigma_{I_s}^2 + I_{s_0}\sum_{i=1}^{s} I_{i_0}(\gamma_i - \alpha_i); \\ &j = 1, 2, \ldots, s-1; \qquad \sum_{i=1}^{s}\gamma_i = 1 \end{aligned}\right\} \qquad \text{(XVIII, 2.19)}$$

sau

$$\left.\begin{aligned} &\gamma_j[\sigma_{I_j}^2 + \sigma_{I_s}^2(I_{j_0} - I_{s_0})^2] + \sum_{i\neq j\neq s}\gamma_i[\sigma_{I_s}^2 + (I_{i_0} - I_{s_0})(I_{j_0} - I_{s_0})] = \\ &\quad = \sigma_{I_s}^2 + \sum_{i\neq s}\alpha_i(I_{i_0} - I_{s_0})(I_{j_0} - I_{s_0}); \\ &j = 1, 2, \ldots, s-1; \quad \gamma_s = 1 - \sum_{i=1}^{s-1}\gamma_i. \end{aligned}\right\} \qquad \text{(XVIII, 2.19')}$$

Dacă $s = 2$, rezolvarea sistemului (XVIII, 2.19) sau (XVIII, 2.19') se face cu ajutorul formulelor:

$$\left.\begin{aligned} \gamma_I &= \frac{\sigma_{I\,II}^2 + \alpha_I(I_{0I} - I_{0II})^2}{\sigma_{I\,I}^2 + \sigma_{I\,II}^2 + (I_{0I} - I_{0II})^2} \\ \gamma_{II} &= \frac{\sigma_{I\,I}^2 + \alpha_{I\,II}^2 + \alpha_{II}(I_{0I} - I_{0II})^2}{\sigma_{I\,I}^2 + \sigma_{I\,II}^2 + (I_{0I} - I_{0\,II})^2} \end{aligned}\right\}. \qquad \text{(XVIII, 2.20)}$$

Folosirea practică a sistemelor (XVIII, 2.19) sau (XVIII, 2.19') sau (XVIII, 2.20) este mai dificilă prin faptul că valorile teoretice I_{i_0} nu ne sînt cunoscute, iar înlocuirea valorilor teoretice I_{i_0} prin cele efective I_i este admisibilă numai în cazul cînd diferența dintre ele ridicată la pătrat este neglijabilă în comparație cu dispersia valorilor teoretice I_{i_0}. În acest caz însă, — el apare atunci cînd toate valorile h_i sînt destul de mari — dispersiile întîmplătoare σ_I^2 din ecuațiile (XVIII, 2.19), (XVIII, 2.19'), (XVIII, 2.20) pot fi neglijate iar sistemul rezolvat cu ajutorul relației:

$$\gamma_i = \alpha_i, \qquad (i = 1, 2, \ldots, s) \qquad \text{(XVIII, 2.21)}$$

ceea ce înseamnă că sistemul de fixare a ponderilor este folosit fără nici o corecție de reprezentativitate.

Ecuațiile (XVIII, 2.19), (XVIII, 2.19'), (XVIII, 2.20) pot fi folosite practic pentru corectarea ponderilor, dacă admitem ipoteze simplificatoare cu privire la valorile teoretice I_{i_0} pe care le putem verifica experimental.

În acest scop, I. I. Lukomski propune *metoda numărului minim de centre de variație.*

Metoda numărului minim de centre de variație. Această metodă poate fi aplicată în două ipoteze:

1°. Să presupunem că toți indicii individuali I_i oscilează la întîmplare în jurul unui centru I_0, adică

$$I_{i_0} = I_0\,, \qquad (i = 1, 2, \ldots, s).$$

După cum se vede, sistemul (XVIII, 2.19) se transformă în sistemul (XVIII, 2.3); sîntem, deci, în cazul cînd la baza fiecărui indice individual stă una și aceeași valoare teoretică:

$$\gamma_j \sigma^2_{I_j} = \gamma_s \sigma^2_{I_s}\,, \qquad (j = 1, 2, \ldots, s-1)$$

sau

$$\Sigma \gamma_i = 1$$

$$\left.\begin{aligned} \frac{\gamma_i}{h_i} &= c \\ \Sigma\gamma_i &= 1 \end{aligned}\right\} \qquad (i = 1, 2, \ldots, s). \qquad \text{(XVIII, 2.22)}$$

Avem de aceea, soluția sistemului:

$$\gamma_i = \frac{h_i}{\Sigma h_i}\,, \qquad (i = 1, 2, \ldots, s). \qquad \text{(XVIII, 2.23)}$$

După cum știm, în acest caz $K_1 < 3$.

Dacă $K_1 > 3$, ipoteza unui singur centru de variație trebuie respinsă, urmînd a se verifica ipoteza cu două centre.

2°. Toți indicii individuali se împart în două grupe, făcîndu-se ipoteza că fiecare din ele își are un centru propriu de variație.

Să presupunem că valorile

$$I_{\mathrm{I}_1}, I_{\mathrm{I}_2}, \ldots, I_{\mathrm{I}_k} \text{ se referă la centrul } I_{\mathrm{I}_0}$$

iar

$$I_{\mathrm{II}_1}, I_{\mathrm{II}_2}, \ldots, I_{\mathrm{II}_l} \text{ se referă la centrul } I_{\mathrm{II}_0}$$

și că $k + l = s$.

Gruparea se efectuează după criteriul mărimii indicilor I_i și al omogeneității materialului la care se referă.

Prin urmare, se consideră că:

$$I_{\mathrm{I}_{10}} = I_{\mathrm{I}_{20}} = I_{\mathrm{I}_{30}} = \ldots = I_{\mathrm{I}_{k_0}} = I_{\mathrm{I}_0}\,,$$
$$I_{\mathrm{II}_{10}} = I_{\mathrm{II}_{20}} = I_{\mathrm{II}_{30}} = \ldots = I_{\mathrm{II}_{l_0}} = I_{\mathrm{II}_0}\,.$$

Scanned with OKEN Scanner

În acest caz relațiile (XVIII, 2.19) pot fi scrise sub forma sistemului de ecuații :

$$\gamma_{I_j}\,\sigma^2_{I_{I_j}} = \gamma_{I_k}\,\sigma^2_{I_{I_k}} \qquad (j = 1, 2, \ldots, k-1) \qquad \text{(XVIII, 2.24a)}$$

$$\gamma_{II_j}\,\sigma^2_{I_{II_j}} = \gamma_{II_l}\,\sigma^2_{I_{II_l}} \qquad (j = 1, 2, \ldots, l-1) \qquad \text{(XVIII, 2.24b)}$$

$$\gamma_{I_k}\,\sigma^2_{I_{I_k}} + I_{i_0}\sum_{i=1}^{s} I_{i_0}(\gamma_i - \alpha_i) = \gamma_{II_l}\sigma^2_{I\,II_l} + I_{II_0}\sum_{i=1}^{s} I_{i_0}(\gamma_i - \alpha_i), \qquad \text{(XVIII, 2.24c)}$$

$$\sum_{i=1}^{s} \gamma_i = 1. \qquad \text{(XVIII, 2.24d)}$$

Din ecuațiile (XVIII, 2.24a) și (XVIII, 2.24b), obținem

$$\frac{\gamma_{Ij}}{h_{I_j}} = C_I\,,$$

în care

$$C_I = \frac{\Sigma_I \gamma_j}{\Sigma_I h_j}\,, \qquad (j = 1, 2, \ldots, k) \qquad \text{(XVIII, 2.25)}$$

și

$$\frac{\gamma_{II_j}}{h_{II_j}} = C_{II}\,,$$

în care

$$C_{II} = \frac{\Sigma_{II}\gamma_j}{\Sigma_{II}h_j}\,, \qquad (j = 1, 2, \ldots, k). \qquad \text{(XVIII, 2.26)}$$

Se vede că *în cadrul fiecărei grupe ponderile γ_j trebuie să fie proporționale cu numerele h_j*.

Să calculăm indicele mediu pentru fiecare grupă, folosind ponderile h_i :

$$I_I = \frac{\Sigma_I I_i h_i}{\Sigma_{II} h_i}\,, \qquad I_{II} = \frac{\Sigma_{II} I_i h_i}{\Sigma_{II} h_i}.$$

Folosind notația

$$\Sigma_I h_i = h_I\,, \qquad \Sigma_{II} h_i = h_{II}\,,$$
$$\Sigma_I \gamma_i = \gamma_I\,, \qquad \Sigma_{II}\gamma_i = \gamma_{II}\,,$$
$$\Sigma_I \alpha_i = \alpha_I\,, \qquad \Sigma_{II}\alpha_i = \alpha_{II}\,,$$

dispersiile indicilor de grupă I_I și I_{II} se calculează cu ajutorul formulei (XVIII, 2.10) :

$$\sigma^2_{I_I} = \frac{2I_0^2 V_0^2}{h_I}\,, \qquad \sigma^2_{I_{II}} = \frac{2I_0^2 V_0^2}{h_{II}}.$$

Ecuația (XVIII, 2.24c) se poate scrie :

$$\gamma_I \sigma^2|_I + I_{I_0}[I_{I_0}(\gamma_I - \alpha_I) + I_{II_0}(\gamma_{II} - \alpha_{II})] =$$
$$= \gamma_{II}\sigma^2|_{II} + I_{II_0}[I_{I_0}(\gamma_I - \alpha_I) + I_{II_0}(\gamma_{II} - \alpha_{II})]. \qquad \text{(XVIII, 2.27)}$$

Am obținut ecuațiile de forma (XVIII, 2.19), în care $s = 2$, însă pentru γ_I și γ_{II} avînd indicii de grupă I_I și I_{II}.

Ecuația (XVIII, 2.27) poate fi adusă la forma (XVIII, 2.19′) și se rezolvă cu ajutorul formulelor (XVIII, 2.20), în care $s = 2$:

$$\left.\begin{aligned}\gamma_I &= \frac{\sigma_{I_I}^2 + \alpha_I + (I_{0I} - I_{0II})^2}{\sigma_{I_I}^2 + \sigma_{I_{II}}^2 + (I_{0I} - I_{0II})^2},\\ \gamma_{II} &= \frac{\sigma_{I_I}^2 + \alpha_{II}(I_{0I} - I_{0II})^2}{\sigma_{I_I}^2 + \sigma_{I_{II}}^2 + (I_{0I} - I_{0II})^2}.\end{aligned}\right\} \qquad \text{(XVIII, 2.28)}$$

Formulele (XVIII, 2.28) sînt identice cu cele (XVIII, 2.20) însă adaptate pentru ponderile celor doi indici de grupă I_I și I_{II}.

Observăm acum că dispersiile $\sigma_{I_I}^2$ și $\sigma_{I_{II}}^2$ ale indicilor de grupă I_I și I_{II}, fundamentate pe valorile totale h_i, pot fi considerate destul de mici în comparație cu dispersiile teoretice $(I_{0I} - I_{0II})^2$. Aceasta ne duce de fapt la primul caz examinat în care ponderile apar fără corecție:

$$\left.\begin{aligned}\gamma_I &= \alpha_I,\\ \gamma_{II} &= \alpha_{II}.\end{aligned}\right\} \qquad \text{(XVIII, 2.29)}$$

Așadar, metoda de calcul al indicelui mediu $\bar{I}$ constă în:

a) calculul indicilor de grupă I_I și I_{II}, folosind ponderile h_i:

$$I_I = \frac{\Sigma_I h_i I_i}{\Sigma_I h_i}, \qquad I_{II} = \frac{\Sigma_{II} h_i I_i}{\Sigma_{II} h_i}; \qquad \text{(XVIII, 2.30)}$$

b) calculul mediei aritmetice a indicilor I_I și I_{II}, folosind ponderile de fixare $\alpha_I = \Sigma_I \alpha_i$ și $\alpha_{II} = \Sigma_{II} \alpha_i$,

$$\bar{I} = I_I \Sigma_I \alpha_i + I_{II} \Sigma_{II} \alpha_i. \qquad \text{(XVIII, 2.31)}$$

Ponderile definitive γ_i ale indicilor individuali I_i din care se formează indicii de grupă I_I și I_{II} capătă următoarea expresie:

$$\gamma_{I_i} = \frac{h_i}{\Sigma_I h_i} \Sigma_I \alpha_i; \qquad \gamma_{II_i} = \frac{h_i}{\Sigma_{II} h_i} \Sigma_{II} \alpha_i. \qquad \text{(XVIII, 2.32)}$$

Aplicarea formulelor (XVIII, 2.30) și (XVIII, 2.31) sau a formulei (XVIII, 2.32) se fundamentează pe ipoteza existenței a două centre de variație.

Verificarea acestei ipoteze se face cu ajutorul criteriului K_2 ce se construiește în mod analog criteriului K_1:

$$K_2 = \left(\frac{D_2}{D_{20}} - 1\right)\sqrt{\frac{s-2}{2}} < 3, \qquad \text{(XVIII, 2.33)}$$

Scanned with OKEN Scanner

în care :

$$D_2 = \frac{\Sigma_I h_i (I_i - I_I)^2 + \Sigma_{II} h_i (I_i - I_{II})^2}{\Sigma_I h_i + \Sigma_{II} h_i}, \qquad \text{(XVIII, 2.34)}$$

$$D_{20} = \frac{2V_0^2 [I_{I0}^2 (s_1 - 1) + I_{II0}^2 (s_2 - 1)]}{\Sigma_I h_i + \Sigma_{II} h_i}. \qquad \text{(XVIII, 2.35)}$$

Formula (XVIII, 2.33) devine astfel :

$$K_2 = \left(\frac{\Sigma_I h_i (I_i - I_I)^2 + \Sigma_{II} h_i (I_i - I_{II})^2}{2V_0^2 [I_{I0}^2 (s_1 - 1) + I_{II0}^2 (s_2 - 1)} - 1 \right) \sqrt{\frac{s-2}{2}} < 3. \qquad \text{(XVIII, 3.33')}$$

Dacă $K_2 > 3$, atunci ipoteza cu două centre de variație trebuie respinsă și înlocuită prin ipoteza cu trei centre etc.

În cazul general, cînd numărul centrelor este m, indicele mediu se calculează în mod analog cazului examinat (cînd $m = 2$), iar criteriul de variabilitate K_m capătă următoarea formă :

$$K_m = \left(\frac{D_m}{D_{m0}} - 1 \right) \sqrt{\frac{s-m}{2}} =$$

$$= \left(\frac{\Sigma_{ij} h_{ij} (I_{ij} - \bar{I}_i)^2}{2V_0^2 \Sigma_i I_{0i}^2 (s_i - 1)} - 1 \right) \sqrt{\frac{s-m}{2}} < 3, \qquad \text{(XVIII, 2.36)}$$

în care :

i este numărul grupei (centrului) ;
j numărul indicelui individual în cadrul grupei ;
I_{0i} este înlocuit cu valorile empirice $\bar{I}_i$ și V_0 ;
s_i numărul indicilor în grupa i ; $s = \Sigma s_i$.

Să considerăm următorul exemplu :

Avem date care caracterizează indicii productivității muncii (I_i) la șase întreprinderi ale unui trust ; de asemenea, date privind numărul muncitorilor din perioadele curentă și de bază la aceleași întreprinderi (vezi tabelul 87).

Tabelul 87

Întreprinderi și grupe de întreprinderi în ipoteza a două și trei centre de variație		Indicii productivității muncii (I_i)	Numărul muncitorilor	
			Perioada de bază (n)	Perioada curentă (n')
I	A	0,30	100	100
	B	0,50	10	400
	C	0,80	100	100
II	D	0,90	10	400
	E	1,30	100	100
	F	1,50	10	400

Ni se cere să calculăm indicele mediu ($\bar{I}$) al productivității muncii pe întregul trust.

Drept ponderi de fixare vor fi luați indicatorii cantitativi privind numărul muncitorilor din perioada curentă (n'), după cum se și procedează de obicei în practică:

$$\alpha_A = \frac{100}{1\,500} = 0{,}066, \qquad \alpha_B = \frac{400}{1\,500} = 0{,}266, \qquad \alpha_C = 0{,}066;$$

$$\alpha_D = 0{,}266, \qquad \alpha_E = 0{,}066, \qquad \alpha_F = 0{,}266.$$

Îndicele mediu calculat pe baza acestor ponderi de fixare va fi

$$\bar{I} = (0{,}30 \cdot 0{,}066) + (0{,}50 \cdot 0{,}266) + (0{,}80 \cdot 0{,}066)\,(0{,}90 \cdot 0{,}266) +$$
$$+ (1{,}30 \cdot 0{,}066) + (1{,}50 \cdot 0{,}266) = 0{,}9298 \text{ sau } 92{,}98\%.$$

După ce vom introduce corecțiile respective după metoda propusă de I. I. Lukomski, vom vedea că acest rezultat este exagerat.

Să calculăm mediile armonice (h_i) pentru fiecare întreprindere:

$$h_A = \frac{2}{\frac{1}{100} + \frac{1}{100}} = 100, \quad h_B = \frac{2}{\frac{1}{10} + \frac{1}{400}} = 19{,}5,$$

$$h_C = \frac{2}{\frac{1}{100} + \frac{1}{100}} = 100, \quad h_D = \frac{2}{\frac{1}{10} + \frac{1}{400}} = 19{,}5,$$

$$h_E = \frac{2}{\frac{1}{100} + \frac{1}{100}} = 100, \quad h_F = \frac{2}{\frac{1}{10} + \frac{1}{400}} = 19{,}5.$$

Să considerăm ipoteza unui singur *centru* de variație, în care caz, indicele mediu al productivității muncii pe ansamblul trustului va fi dat de raportul:

$$\bar{I} = \frac{\Sigma h_i I_i}{\Sigma h_i} = \frac{(100 \cdot 0{,}30) + (19{,}5 \cdot 0{,}50) + (100 \cdot 0{,}80) + (19{,}5 \cdot 0{,}90) +}{358{,}5}$$

$$\frac{+ (100 \cdot 1{,}30) + (19{,}5 \cdot 1{,}50)}{358{,}5} = \frac{296}{358{,}5} = 0{,}8256 \text{ sau } 82{,}56\%.$$

Să verificăm dacă a fost cazul să considerăm această ipoteză, calculînd în acest scop criteriul de variabilitate K_1 pentru a vedea dacă este mai mic decît numărul 3 (vom presupune că V_0 este aproximativ egal cu 0,3).

Aplicînd formula (XVIII, 2.16′) vom obține

$$K_1 = \left[\frac{100\cdot(0{,}30-0{,}8256)^2+19{,}5\,(0{,}50-0{,}8256)^2+100\,(0{,}80-0{,}8256)^2+}{2\cdot 0{,}8256^2\cdot 0{,}3^2\cdot(6-1)}\right.$$

$$\left.\frac{+19{,}5(0{,}90-0{,}8256)^2+100(1{,}30-0{,}8256)^2+19{,}5(1{,}50-0{,}8256)^2}{2\cdot 0{,}8256^2\cdot 0{,}3^2\cdot(6-1)}-1\right]\cdot$$

$$\cdot\sqrt{\frac{6-1}{2}}=\left(\frac{60{,}40}{0{,}61}-1\right)\cdot\sqrt{\frac{6-1}{2}}=97{,}3\cdot 1{,}58\cong 153{,}7.$$

După cum se vede, inegalitatea $K_1 = 153{,}7 > 3$ ne spune că ipoteza unui singur centru de variație trebuie respinsă, ca nefiind conformă realității.

Să considerăm acum ipoteza a *două centre de variație*, grupînd indicii individuali în două grupe: grupa I, întrunind indicii productivității muncii de la întreprinderile A, B și C și grupa II, întrunind indicii productivității muncii de la întreprinderile D, E și F.

Vom calcula, în primul rînd, indicii productivității muncii pe aceste două grupe, aplicînd formulele (XVIII, 2.30):

$$\bar{I}_{\mathrm{I}}=\frac{\Sigma_{\mathrm{I}}h_iI_i}{\Sigma_{\mathrm{I}}h_i}=\frac{(0{,}30\cdot 100)+(0{,}50\cdot 19{,}5)+(0{,}80\cdot 100)}{100+19{,}5+100}=\frac{119{,}75}{219{,}5}=0{,}545,$$

$$\bar{I}_{\mathrm{II}}=\frac{\Sigma_{\mathrm{II}}h_iI_i}{\Sigma_{\mathrm{II}}h_i}=\frac{(0{,}90\cdot 19{,}5)+(1{,}30\cdot 100)+(1{,}50\cdot 19{,}5)}{19{,}5+100+19{,}5}=\frac{176}{139}=1{,}26.$$

În al doilea rînd, vom calcula indicele mediu pe ansamblul trustului după formula (XVIII, 2.31):

$$\bar{I}_{\mathrm{I+II}}=I_{\mathrm{I}}\alpha_i+I_{\mathrm{II}}\alpha_i.$$

Avem

$$\alpha_{\mathrm{I}}=\alpha_A+\alpha_B+\alpha_C=0{,}066+0{,}266+0{,}066=0{,}398,$$

$$\alpha_{\mathrm{II}}=\alpha_D+\alpha_E+\alpha_F=0{,}266+0{,}066+0{,}266=0{,}598$$

și atunci

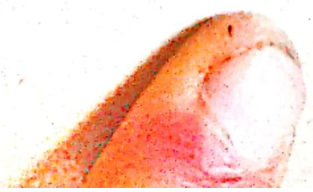
$$\bar{I}_{\mathrm{I+II}}=(0{,}545\cdot 0{,}398)+(1{,}26\cdot 0{,}598)\cong 0{,}96 \text{ sau } 96\,\%.$$

Să verificăm justețea ipotezei cu două centre de variație, calculînd criteriul de variabilitate K_2, după formula (XVIII, 2.33′):

$$K_2=\left[\frac{100\cdot(0{,}30-0{,}545)^2+19{,}5\,(0{,}50-0{,}545)^2+100\cdot(0{,}80-0{,}545)^2+}{2\cdot 0{,}3^2\,[0{,}545^2\cdot(2-1)+1{,}26^2\cdot(2-1)]}\right.$$

$$\left.\frac{+19{,}5\cdot(0{,}90-1{,}26)^2+100\,(1{,}30-1{,}26)^2+19{,}5\cdot(1{,}50-1{,}26)^2}{2\cdot 0{,}3^2\,[0{,}545^2\cdot(2-1)+1{,}26^2\cdot(2-1)]}-1\right]\cdot$$

$$\cdot\sqrt{\frac{6-2}{2}}=\left[\frac{16{,}208}{0{,}33}-1\right]\cdot\sqrt{2}=(49{,}09-1)\cdot 1{,}41\cong 67{,}68.$$



Prin urmare, $K_2 = 67{,}68 > 3$, ceea ce înseamnă că nici ipoteza cu două centre de variații nu este conformă realității și de aceea trebuie respinsă.

În sfîrșit, să considerăm și ipoteza cu trei centre de variație, grupînd indicii individuali în trei grupe:

— grupa I, întrunind indicii productivității muncii din întreprinderile A și B;

— grupa II, întrunind indicii productivității muncii din întreprinderile C și D și

— grupa III, întrunind indicii productivității muncii din întreprinderile E și F.

Indicii medii de grupă vor fi:

$$\bar{I}_{\text{I}} = \frac{\Sigma_{\text{I}} h_i I_i}{\Sigma_{\text{I}} h_i} = \frac{(0{,}30 \cdot 100) + (0{,}50 \cdot 19{,}5)}{119{,}5} = \frac{39{,}75}{119{,}5} = 0{,}33,$$

$$\bar{I}_{\text{II}} = \frac{\Sigma_{\text{II}} h_i I_i}{\Sigma_{\text{II}} h_i} = \frac{(0{,}80 \cdot 100) + (0{,}90 \cdot 19{,}5)}{119{,}5} = \frac{97{,}55}{119{,}5} = 0{,}81,$$

$$\bar{I}_{\text{III}} = \frac{\Sigma_{\text{III}} h_i I_i}{\Sigma_{\text{III}} h_i} = \frac{(1{,}30 \cdot 100) + (1{,}50 \cdot 19{,}5)}{119{,}5} = \frac{159{,}25}{119{,}5} 1{,}33,$$

$$\bar{I}_{\text{I+II+III}} = \bar{I}_{\text{I}}\, \alpha_{\text{I}} + \bar{I}_{\text{II}}\, \alpha_{\text{II}} + \bar{I}_{\text{III}}\, \alpha_i .$$

Avem

$$\alpha_{\text{I}} = \alpha_A + \alpha_B = 0{,}332,$$
$$\alpha_{\text{II}} = \alpha_C + \alpha_D = 0{,}332,$$
$$\alpha_{\text{III}} = \alpha_E + \alpha_F = 0{,}332.$$

Atunci

$$\bar{I}_{\text{I+II+III}} = (0{,}33 \cdot 0{,}332) + (0{,}81 \cdot 0{,}332) + (1{,}33 \cdot 0{,}332) =$$
$$= 0{,}1095 + 0{,}269 + 0{,}4415 = 0{,}81 \text{ sau } 81\%.$$

Să verificăm și de data aceasta justețea ipotezei cu trei centre de variație, calculînd criteriul de variabilitate K_3, după formula (XVIII, 2.36):

$$K_3 = \left[\frac{100 \cdot (0{,}30 - 0{,}33)^2 + 19{,}5\,(0{,}50 - 0{,}33)^2 + 100\,(0{,}80 - 0{,}81)^2 + 19{,}5(0{,}90 - 0{,}81^2) + 100(1{,}30 - 1{,}33)^2 + 19{,}5(1{,}50 - 1{,}33)^2}{2.0{,}3^2\,[0{,}33^2 \cdot (3-1) + 0{,}81^2\,(3-1) + 1{,}33^2\,(3-1)]} - 1 \right] \cdot \sqrt{\frac{6-3}{2}} =$$

$$= \left(\frac{1{,}46}{1{,}38} - 1 \right) \sqrt{\frac{3}{2}} \cong 0{,}07.$$

De data aceasta observăm că $K_3 = 0{,}07 < 3$, ceea ce înseamnă în exemplul nostru că numărul minim al centrelor de variație trebuie să fie egal cu trei.

Examinînd cu atenție exemplul ales, putem constata că într-adevăr cea mai rațională grupare a celor șase indici ai productivității muncii

este cea care-i desparte în trei grupe omogene după mărimea nivelului lor, după cum am procedat alegînd ipoteza cu trei centre de variație.

Rezultă deci că indicele productivității muncii pe ansamblul trustului este egal cu 0,81 sau 81% și nu 92%, învederînd o scădere a productivității muncii cu 13%. Acesta este rezultatul real care s-a dedus prin eliminarea factorilor întîmplători, considerîndu-se acțiunea legii numerelor mari.

*

Cu privire la sistemul elaborat de I. I. Lukomski avem de făcut următoarele observații:

1) Studiul lui I. Lukomski aduce contribuții importante în teoria indicilor statistici. Ideea sa de a ține seama de numărul unităților din colectivitate la construirea unui indice și de a măsura valabilitatea concluziilor care se pot trage din examinarea indicelui în raport cu numărul unităților este fundamentală în statistică. Pare de aceea curios cum n-a fost aplicată pînă acum la studiul indicilor.

Să considerăm un atelier care are 100 de lucrători în perioada de bază și 100 lucrători în perioada curentă. Oricine va spune că procedăm greșit dacă raportăm productivitatea muncii tuturor lucrătorilor din perioada curentă la productivitatea muncii unui singur lucrător din perioada de bază. Dacă acest unic lucrător are productivitatea mare, vom obține pentru productivitatea muncii din perioada curentă o valoare mică și invers, dacă lucrătorul luat etalon de măsură în perioada de bază, luat la întîmplare, are productivitatea mică, vom ajunge la concluzia că productivitatea muncii în perioada curentă a crescut foarte mult. Obținem astfel rezultate diferite care depind de productivitatea muncii lucrătorului luat ca etalon din perioada de bază, care nu ne poate duce la nici o concluzie temeinică.

Dacă vom lua la întîmplare 10 lucrări din perioada de bază, adică o selecție de mărimea 10%, și în raport cu productivitatea medie a muncii lor vom măsura productivitatea muncii din perioada curentă, este evident că vom obține un rezultat mai apropiat de adevăr și cu cît vom considera un număr mai mare de lucrători în perioada de bază cu atît vom lucra în condiții mai aproape de realitate.

Legea numerelor mari joacă un rol și mai important la seriile în timp pentru care variațiile accidentale sînt inevitabile *).

---

*) În această categorie intră de exemplu toate performanțele sportive. Fiecare atlet are la un moment dat o performanță medie în jurul căreia se grupează toate realizările lui. Un aruncător cu greutatea nu posedă precizia de a obține rezultate riguros egale la toate aruncările.

Să admitem, ținînd seama de observația de mai sus, că un grup de atleți este supus metodic unui antrenament la aruncarea greutății. Cum putem ști dacă la sfîrșitul antrenamentului s-a realizat un progres efectiv? Este natural, ca la începutul antrenamentului să facem o statistică a aruncărilor realizate de atleții care fac parte din grup. Să admitem că s-a obținut o medie generală m. După terminarea antrenamentului aceeași statistică ne conduce la media generală m'. Dacă $m' > m$ putem deduce că s-a realizat un progres? Răspunsul nu e întotdeauna afirmativ. Natura statistică a acestei serii de date ne împiedică să dăm un răspuns înaintea efectuării unei analize mai adînci. Mai întîi numerele m și m' nu sînt fixe, deoarece, dacă un atlet aruncă de 5 ori greutatea obține cinci rezultate diferite, este evident că și numerele m și

2) Formula (XVIII, 2.18) nu reprezintă dispersia indicelui

$$\bar{I}_0 = \Sigma \gamma_i I_i .$$

O teoremă generală din statistică ne dă dispersia unei sume liniare de variabile independente. În baza acestei teoreme, avem

$$D^2(I_0) = \Sigma \gamma_i^2 D^2(I_i)$$

adică tocmai formula care ne-a condus la rezultatele din cazul A (p. 480).

3) Considerațiile de mai sus prezintă un interes deosebit cînd un indice general I rezultă din ponderarea unor indici parțiali $I_1, I_2, \ldots, I_s$, adică atunci cînd

$$I = \alpha_1 I_1 + \alpha_2 I_2 + \ldots + \alpha_s I_s$$

unde

$$\alpha_i \geqslant 0, \quad \sum_{i=1}^{s} \alpha_i = 1.$$

Exactitatea lui I depinde de gradul de exactitate al indicilor individuali $I_1, I_2, \ldots, I_s$. De aceea va trebui să dăm o pondere mai mare indicilor individuali pentru care gradul de exactitate este mai mare și o pondere mai mică indicilor individuali cu un grad de exactitate mai mic. Gradul de exactitate este dat de dispersia indicelui, întrucît intervalul probabil de variație al unui indice este direct proporțional cu dispersia lui.

Întrebuințînd notațiile lui I. Lukomski, avem

$$I_k = \frac{\bar{x}_k'}{\bar{x}_k},$$

$\bar{I}_k$ fiind un indice individual al productivității muncii într-un sector.

Avem

$$D^2(I_k) = \left[\frac{M(\bar{x}_k')}{M(\bar{x}_k)}\right]^2 [V_k^2 + V_k'^2],$$

unde însemnăm prin $M(\bar{x}_k)$ și $M(\bar{x}_k')$ respectiv mediile teoretice ale lui $\bar{x}_k$ și $\bar{x}_k$, iar prin V_k și V_k' coeficienții de variație ai lui $\bar{x}_k$ și $\bar{x}_k'$. Această formulă este stabilită în ipoteza că între $\bar{x}_k$ și $\bar{x}_k'$ există o corelație nulă, ceea ce evident putem să admitem.

Să notăm pentru simplificare:

$$M(\bar{x}_k) = M_k, \quad M(\bar{x}_k') = M_k'.$$

---

m' variază de la probă la probă. De exemplu, dacă $m = 14{,}30$ și $m' = 14{,}45$, avem $m' > m$, dar nu se poate afirma că s-a realizat un progres efectiv, deoarece diferența dintre m' și m, egală cu 0,15 m ar putea să nu fie semnificativă. Va trebuie în acest caz să aplicăm criteriul t al lui Student. Pe baza acestui criteriu vom stabili dacă diferența $m'-m$ este semnificativă sau nu și deci dacă se poate vorbi despre un progres. În această analiză, legea numerelor mari joacă un rol de seamă pentru că cu cît numărul probelor este mai mare cu atît lucrăm în condiții mai mari de exactitate.

Exemplul de mai sus iese din cadrul economiei unde, evident, influențele accidentale sînt neînsemnate.

Formula de mai sus se poate scrie :

$$D^2(I_k) = \frac{M_k'^2}{M_k^2}[V_k^2 + V_k'^2].$$

Admitem că pentru toți lucrătorii din perioada de bază există aceeași dispersie σ_k^2, iar pentru toți lucrătorii din perioada curentă dispersia $\sigma_k'^2$.

Expresia de mai sus devine

$$D^2(I_k) = \frac{M_k'^2}{M_k^2}\left[\frac{\sigma_k^2}{n_k M_k^2} + \frac{\sigma_k'^2}{n_k' M_k'^2}\right],$$

unde n_k și n_k' reprezintă numărul lucrătorilor din perioada de bază și perioada curentă.

Se vede din această formulă că cu cît n_k și n_k' sînt mai mari cu atît dispersia este mai mică. Dacă n_k este foarte mic, de exemplu $n_k = 1$, $D^2(I_k)$ este foarte mare. Indicele I_k pierde în acest caz orice semnificație.

4) Admițînd că există s sectoare, este normal să încercăm să determinăm indicele mediu al productivității muncii pentru toate sectoarele printr-o metodă diferită de a lui Lukomski.

Vom considera ca indice mediu al productivității muncii expresia

$$\bar{I} = \frac{\lambda_1 \bar{x}_1^1 + \lambda_2 \bar{x}_2^1 + \ldots + \lambda_s \bar{x}_s^1}{\lambda_1 \bar{x}_1 + \lambda_2 \bar{x}_2 + \ldots + \lambda_s \bar{x}_s}, \qquad (\alpha)$$

unde $\lambda_i \geqslant 0$ sînt anumite ponderi.

Această formulare a indicelui mediu al productivității muncii este cel mai mult folosită în literatura de specialitate. În felul acesta punem la numărător o expresie care măsoară, printr-un sistem de ponderare, productivitatea medie a muncii în perioada curentă, iar la numitor o expresie, care măsoară, cu ajutorul aceluiași sistem de ponderare, productivitatea medie a muncii în perioada de bază.

Aplicînd aceeași metodă, obținem

$$D^2(I) = \frac{(\Sigma \lambda_i' M_i')^2}{(\Sigma \lambda_i M_i)^2}\left[\frac{\Sigma \lambda_i'^2 D^2(\bar{x}_i')}{(\Sigma \lambda_i' M_i')^2} + \frac{\Sigma \lambda_i^2 D^2(\bar{x}_i)}{(\Sigma \lambda_i M_i)^2}\right],$$

unde notațiile sînt aceleași ca mai sus.

Pentru stabilirea acestei formule admitem că covarianta variabilelor x_i' și x_j este nulă, oricare ar fi indicii i și j.

Găsirea minimului lui $D^2(I)$ conduce la calcule foarte complicate. Fără ipoteze suplimentare, a căror valabilitate este discutabilă, aceste calcule nu pot conduce la soluții acceptabile.

Sîntem deci constrînși să obținem indicele mediu al productivității muncii pentru toate sectoarele prin aplicarea unei formule de tipul :

$$I = \lambda_1 I_1 + \lambda_2 I_2 + \ldots + \lambda_s I_s, \qquad (\beta)$$

unde

$$\lambda_i \geqslant 0, \quad \sum_1^s \lambda_i = 1.$$

Observația de mai sus pune în evidență faptul că trebuie să dăm o preferință formulei (β) față de (α) numai din necesități de calcul. În realitate, formula (α) este cea mai indicată pentru că este natural să determinăm un indice al productivității muncii raportînd producția medie din perioada curentă la producția medie din perioada de bază.

Mai avem de observat că chiar aplicarea formulei (β) nu conduce la egalarea ponderilor λ_i cu numere proporționale cu media armonică a numerelor n_i și n_i' decît dacă admitem următoarele ipoteze:

$$M(I_i) = M\left(\frac{\bar{x}_i'}{\bar{x}_i}\right) = I_0, \qquad (i = 1, 2, \ldots, s),$$

$$\frac{D^2(x_i')}{M(\bar{x}_i')} = \frac{D^2(x_i)}{M(\bar{x}_i)} = k_0, \qquad (i = 1, 2, \ldots, s).$$

Prima condiție este foarte restrictivă. Ea exprimă faptul că indicele productivității muncii pentru fiecare sector este aproximativ același. Este evident atunci că și indicele total al productivității muncii este egal cu valoarea comună a indicilor individuali ai productivității muncii. Pe noi ne interesează însă stabilirea unei formule tocmai în cazul cînd indicii productivității muncii variază de la sector la sector. Se ajunge, prin urmare, la o ponderare proporțională cu media armonică a numărului de lucrători numai în cazul cînd nu există nici o dificultate pentru un calcul direct al indicelui total al productivității muncii. Problema rămîne însă deschisă în cazul general.

Condiția a doua poate fi admisă, fără a ne depărta prea mult de realitate, însă nu este suficientă pentru rezolvarea problemei.

5) Concluzia la care ajungem este că problema pusă de I. Lukomski este interesantă, dar nu este rezolvată decît în anumite condiții restrictive.

Dacă aplicăm oricare dintre formulele (α) sau (β) este absolut necesar ca ponderile λ_i să fie funcție de n_i și n_i', adică de numărul lucrărilor în perioada de bază și în perioada curentă

$$\lambda_i = f_i(n_i, n_i').$$

Această funcție trebuie să fie pozitivă și crescătoare în raport cu numerele n_i și n_i'. Pentru valorile mici ale lui n_i sau n_i' funcția trebuie să ia valori mici.

Media armonică îndeplinește această condiție.

Într-adevăr, dacă luăm

$$\lambda_i = \frac{k}{\frac{1}{n_i} + \frac{1}{n_i'}},$$

unde k este o constantă oarecare și presupunem că

$$n_i + n_i' = 2A,$$

unde A este de asemenea o constantă, observăm că λ_i ia valoarea maximă pentru $n_i = n_i' = A$, cînd avem $\lambda_i = \frac{kA}{2}$. Cea mai mică valoare este luată pentru $n_i = 1$, $n_i' = 2A - 1$ sau $n_i = 2A - 1$, $n_i' = 1$, cînd $\lambda_i = k - \frac{k}{2A}$.

Dar în afară de media armonică mai există şi alte medii care se bucură de această proprietate, ca de exemplu media geometrică.

Va fi interesant de stabilit, dintre toate formulele care îndeplinesc cerinţele mai sus arătate, acele formule care rezultă în mod riguros pe baza unui principiu de reprezentativitate, aşa cum a făcut I. I. Lukomski în anumite condiţii restrictive.

Analiza pe care am făcut-o arată că problema pusă este departe de a putea fi considerată ca fiind rezolvată. Studiul lui I. I. Lukomski are însă meritul de a iniţia o serie întreagă de cercetări, care vor duce la mari progrese în teoria indicilor.

Strîns legată de ideea lui Lukomski de a folosi distribuţiile teoretice la analiza statistică este problema pe care o tratăm mai jos. Cercetarea matematică este însă de data aceasta îndreptată pe o cale cu totul diferită, fiind vorba de a se putea determina pragul de semnificaţie al unor indici cu ajutorul metodelor statisticii matematice.

Într-un atelier lucrează pe perioada de timp T_0 un număr de n_0 maşini identice, complet automatizate. Pentru precizare, vom considera n_0 maşini automate de ţesut. În timpul producţiei maşinile au anumite întreruperi, datorită ruperilor de fire, dereglărilor temporare, diverselor calităţi de fire etc. Să notăm prin $x_1, x_2, \ldots, x_{n_0}$ producţia respectivă a celor n_0 maşini pe perioada T_0 şi prin P_{n_0} producţia medie

$$P_{n_0} = \frac{x_1 + x_2 + \ldots + x_{n_0}}{n_0}.$$

Dacă în acelaşi atelier în perioada de timp T_1, lucrează un grup omogen de n_1 maşini în aceleaşi condiţii, astfel încît diferenţa de randament nu poate fi datorată decît unor cauze întîmplătoare, situaţia se prezintă ca şi cum am avea o variabilă aleatoare Y căreia îi cunoaştem n_1 realizări $y_1, y_2, \ldots, y_{n_1}$.

Producţia medie în noua perioadă de timp T_1 este:

$$P_{n_1} = \frac{y_1 + y_2 + \ldots + y_{n_1}}{n_1}.$$

În sfîrşit, dacă la timpul T_2 lucrează un grup de n_2 maşini şi notăm variabila aleatoare corespunzătoare cu z iar cele n_2 realizări cu $z_1; z_2; \ldots; z_{n_2}$ producţia medie în timpul T_2 este

$$P_{n_2} = \frac{z_1 + z_2 + \ldots + z_{n_2}}{n_2}.$$

Dacă considerăm ca perioadă de bază timpul T_0, indicele de productivitate corespunzător perioadei T_1 este $i_1 = \frac{P_{n_1}}{P_{n_0}}$ iar cel corespunzător perioadei T_2 este $i_2 = \frac{P_{n_2}}{P_{n_0}}$.

Problema pe care ne-o punem acum este următoarea:

Se calculează indicii i_1 și i_2 corespunzători celor două perioade și se cere să se indice criterii corespunzătoare unui prag de siguranță dat care să permită a se distinge cazurile în care diferența dintre i_1 și i_2 se datorește unor cauze diferite de acelea în care diferența este semnificativă.

Cu alte cuvinte, dacă din considerarea indicilor i_1 și i_2 putem trage concluzia că mașinile lucrează în aceleași condiții sau că între timp s-a produs un deranjament în funcționarea lor *).

Vom examina mai întîi cazul cînd n_0, n_1, n_2 sînt suficient de mari pentru ca din punct de vedere teoretic să putem considera distribuțiile asimptotice ale indicilor considerați.

În primul rînd este de observat că întrucît ceea ce interesează sînt indicii i_1 și i_2, nu e nevoie chiar de mărimile absolute ale lui x_i, y_j, z_k ci numai de valori relative de forma:

$$x_i' = \frac{x_i}{A}, \quad y_j' = \frac{y_j}{A}, \quad z_k' = \frac{z_k}{A},$$

unde A e un număr ales convenabil (de exemplu A este măsura normei de lucru la timpul T_0).

Dacă se notează:

$$M[x_i'] = a_0; \quad M[y_j'] = a_1; \quad M(z_k') = a_2$$

și cu:

$$A_0 = \max |x_i' - a_0|; \quad A_1 = \max |y_j' - a_1|; \quad A_2 = \max |z_k' - a_2|$$
$$i = 1,2,\ldots,n_0 \qquad j = 1,2,\ldots,n_1 \qquad k = 1,2,\ldots,n_2$$

desigur A_0, A_1, A_2 vor fi numere determinate comparabile cu a_0, a_1, a_2 deci astfel ca

$$\frac{A_0}{\sqrt{n_0}} \to 0; \quad \frac{A_1}{\sqrt{n_1}} \to 0; \quad \frac{A_2}{\sqrt{n_2}} \to 0$$

cînd $n_0, n_1, n_2 \to \infty$.

Presupunînd că x_1', x_2', x_{n_0}' sînt mărimi independente și că $M[x_i' - a_0]^2 = \sigma_0^2$, atunci pe baza teoremei lui Liapunov mărimea aleatoare

$$X_{n_0} = \frac{x_1' + x_2' + \ldots + x_{n_0}'}{n_0}$$

este asimptotic normală cu media a_0 și dispersia $\frac{\sigma_0}{\sqrt{n_0}}$. Pe baza teoremei lui Cebîșev, X_{n_0} tinde în probabilitate către a_0.

Pe baza aceleiași teoreme, mărimea aleatoare

$$T_{n_1;n_2} = Y_{n_1} - Z_{n_2} = \frac{y_1' + y_2' + \ldots + y_{n_1}'}{n_1} - \frac{z_1' + z_2' + \ldots + z_{n_2}'}{n_2}$$

unde se presupune că $y_1', y_2', \ldots, y_{n_1}'$; $z_1', z_2', \ldots, z_{n_2}'$ sînt mărimi independente și $M[y_j' - a_1]^2 = \sigma_1^2$ și $M[z_k' - a_2]^2 = \sigma_2^2$, este asimptotic normală cu media $(a_1 - a_2)$ și dispersia $\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$.

Faptul că am considerat $n_0, n_1, n_2 \to \infty$ ne-a scutit de orice ipoteză asupra legii de distribuție a variabilelor aleatoare X, Y, Z.

Trecem acum la problema noastră care constă în a găsi condiția ca diferența dintre indicii i_1 și i_2 să nu fie semnificativă deci să fie datorată numai influențelor aleatoare.

*) Problema a fost tratată în lucrarea Gh. Mihoc și G. Theiler, *Legile de probabilitate ale unor indici statistici*. Analele Universității „C. I. Parhon", XXIV (1961).

Cum

$$i_1 - i_2 = \frac{Y_{n_1}}{X_{n_0}} - \frac{Z_{n_2}}{X_{n_0}} = \frac{Y_{n_1} - Z_{n_2}}{X_{n_0}}$$

dacă presupunem că Y și Z sînt independente, pe baza celor arătate anterior:

$$M[Y_{n_1} - Z_{n_2}] = a_1 - a_2;\ D^2(Y_{n_1} - Z_{n_2}) = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}.$$

În cazul cînd situația în timpul T_2 nu a suferit modificări față de timpul T_1 exceptînd din cauze întîmplătoare, atunci $a_1 = a_2$ și dacă și $\sigma_1 = \sigma_2 = \sigma$, atunci

$$M[Y_{n_1} - Z_{n_2}] = 0;\quad D[Y_{n_1} - Z_{n_2}] = \sigma\sqrt{\frac{n_1+n_2}{n_1 n_2}}.$$

Vom folosi acum proprietatea constînd în faptul că dacă x_n și y_n sînt asimptotic normale $\left(a, \frac{h}{\sqrt{n}}\right)$ și $\left(b, \frac{k}{\sqrt{n}}\right)$, $b \neq 0$, mărimea

$$Z_n = \sqrt{n}\,\frac{x_n - a}{y_n}$$

este asimptotic normală $\left(0, \frac{h}{b}\right)$, a cărei demonstrație nu o dăm aici.

Mărimea aleatoare:

$$\sqrt{\frac{n_1 n_2}{n_1 + n_2}}\,(i_1 - i_2) = \sqrt{\frac{n_1 n_2}{n_1 + n_2}}\,\frac{Y_{n_1} - Z_{n_2}}{X_{n_0}}$$

este asimptotic normală cu media 0 și dispersia $\frac{\sigma}{a_0}$ în baza proprietății menționate.

Aceasta înseamnă că asimptotic probabilitatea inegalității:

$$t_1\frac{\sigma}{a_0}\sqrt{\frac{n_1+n_2}{n_1 n_2}} < i_1 - i_2 < t_2\frac{\sigma}{a_0}\sqrt{\frac{n_1+n_2}{n_1 n_2}}$$

este

$$\frac{1}{\sqrt{2\pi}}\int_{t_1}^{t_2} e^{-\frac{x^2}{2}}\,dx.$$

Dacă se ia $t_1 = -3$, $t_2 = 3$, valoarea asimptotică a probabilității este 0,9973. Am obținut astfel un criteriu pentru a răspunde la problema propusă.

Desigur a_0 și σ nu sînt cunoscute. Știm însă din teoria estimației că

$$\bar{x} = \frac{x_1 + x_2 + \ldots + x_{n_0}}{n_0} \text{ și } s^2 = \frac{\sum_1^{n_1}(y_j - \bar{y})^2 + \sum_1^{n_2}(z_k - \bar{z})^2}{n_1 + n_2 - 2},$$

sînt estimații nedeplasate cu dispersia minimă pentru a_0 și σ^2

Pentru σ s-ar putea lua estimația corectă

$$s = \sqrt{\frac{\sum_1^{n_1}(y_j - \bar{y})^2 + \sum_1^{n_2}(z_k - \bar{z})^2}{n_1 + n_2 - 2}}.$$

Există pentru σ și estimații nedeplasate cu dispersia tinzînd la 0 cînd $n \to \infty$. (A se vedea O. Onicescu și Gh. Mihoc, *Lecții de statistică matematică*, Ed. tehnică, București, p. 544), dar utilizarea lor practică e mai dificilă.

Mai rămîne de cercetat legea de distribuție a mărimii aleatoare $\frac{s}{\bar{x}}$, studiu care nu prezintă dificultăți speciale, dar pe care nu-l vom face aici.

Cunoscînd această lege se poate calcula:

$$M\left[\frac{s}{\bar{x}}\right] \text{ și } M\left[\frac{s}{\bar{x}} - M\left(\frac{s}{\bar{x}}\right)\right]^2$$

și apoi stabili regiunile de încredere pentru un prag de siguranță dat.

Pentru a stabili variația posibilă a diferenței $i_1 - i_2$ am presupus că variabilele Y și Z au nu numai aceiași valoare medie, ci și aceeași dispersie. Or, dacă ipoteza $a_1 = a_2$ este evident necesară pentru considerațiile de mai sus, ipoteza $\sigma_1 = \sigma_2$ nu este necesară.

De aceea vom considera acum cazul în care $\sigma_1 \neq \sigma_2$ deci situația în care de la timpul T_1 la T_2 singura schimbare care a intervenit a fost o creștere sau descreștere a influenței factorilor aleatori.

În acest caz

$$M[Y_{n_1} - Z_{n_2}] = 0; \; D[Y_{n_1} - Z_{n_2}] = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}.$$

Vom considera aceeași mărime aleatoare $\sqrt{\frac{n_1 n_2}{n_1 + n_2}}\,(i_1 - i_2)$ care va fi tot asimptotic normală cu media 0 și dispersia

$$\lim_{\frac{n_1 n_2}{n_1+n_2} \to \infty} \sqrt{\frac{n_1 n_2}{n_1 + n_2}} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}.$$

Dar aici spre diferență de cazul precedent nu este suficient să presupunem că

$$\frac{n_1 n_2}{n_1 + n_2} \to \infty.$$

Trebuie presupus în plus că $\frac{n_2}{n_1}$ tinde către o limită determinată cînd n_1 respectiv $n_2 \to \infty$.

În acest caz notînd

$$\lim \frac{n_2}{n_1 + n_2} = k,$$

se vede că legea normală limită are media 0 și dispersia $\sqrt{k\sigma_1^2 + (1-k)\sigma_2^2}$, iar de aici se deduce imediat că asimptotic, probabilitatea inegalității:

$$t_1 \sqrt{\frac{k\sigma_1^2 + (1-k)\sigma_2^2}{a_0}} \sqrt{\frac{n_1 + n_2}{n_1 n_2}} < i_1 - i_2 < t_2 \sqrt{\frac{k\sigma_1^2 + (1-k)\sigma_2^2}{a_0}} \sqrt{\frac{n_1 + n_2}{n_1 n_2}}$$

este

$$\frac{1}{\sqrt{2\pi}} \int_{t_1}^{t_2} e^{-\frac{x^2}{2}}\, dx.$$

Desigur că și aici în loc de a_0 se va lua $\bar{x}$ iar în loc de σ_1^2 și σ_2^2:

$$s_1^2 = \frac{\sum_1^{n_1} (y_j - \bar{y})^2}{n_1 - 1}\ ;\ s_2^2 = \frac{\sum_1^{n_2} (z_k - \bar{z})^2}{n_2 - 1}.$$

Se poate studia legea de distribuție a mărimii aleatoare

$$\sqrt{\left(1 - \frac{1}{n_1}\right) s_1^2 + \left(1 - \frac{1}{n_2}\right) s_2^2} : \bar{x}$$

spre a se putea aprecia cu mai mare precizie gradul de aproximație obținut prin substituțiile de mai sus, studiu pe care nu-l mai întreprindem aici.

Pînă acum am considerat o singură categorie de mașini. Nimic nu ne împiedică să considerăm un număr determinat s de categorii $X^{(1)}, X^{(2)}, \ldots, X^{(s)}$ cu un număr de realizări $n_0^{(1)}, n_0^{(2)}, \ldots, n_0^{(s)}$ în perioada T_0 apoi variabilele aleatoare $Y^{(1)}, Y^{(2)}, \ldots, Y^{(s)}$ cu un număr de realizări $n_1^{(1)}, n_1^{(2)}, \ldots, n_1^{(s)}$ în perioda T_1 și variabilele aleatoare $Z^{(1)}, Z^{(2)}, \ldots, Z^{(s)}$ cu un număr de realizări $n_2^{(1)}, n_2^{(2)}, \ldots, n_2^{(s)}$ în perioada T_2.

Dacă se presupune că $X^{(i)}, Y^{(s)}, Z^{(k)}$ sînt independente $i, j, k = 1, 2, \ldots, s$, și de asemenea fiecare realizare este independentă de celelalte, problema se încadrează într-un tip analog celui de mai sus cu condiția ca $n_0^{(i)}, n_1^{(s)}, n_2^{(k)}$ să tindă la ∞ cu o anumită regularitate.

Acum ne vom ocupa de următoarea problemă.

Fiind dați doi indici de productivitate corespunzători mașinilor din două ateliere diferite, cînd se poate spune că aceștia diferă semnificativ?

Cu alte cuvinte dacă la atelierul I producția medie a fost la timpul T_1 de P_1^{I} și la timpul T_2 de P_2^{I}, iar la atelierul II de P_1^{II} respectiv P_2^{II}, cînd putem spune că

$$i_{\mathrm{I}} = \frac{P_2^{\mathrm{I}}}{P_1^{\mathrm{I}}} \quad \text{și} \quad i_{\mathrm{II}} = \frac{P_2^{\mathrm{II}}}{P_1^{\mathrm{II}}}$$

diferă semnificativ și cînd diferența poate fi atribuită factorilor întîmplători?

Vom începe și aici prin a considera un caz mai simplu care matematic poate fi prezentat astfel:

Fie variabilele aleatoare X^{I} și Y^{I} asupra cărora se fac n_1 respectiv m_1 experiențe independente în perioada T_1, rezultatele acestora fiind $x_1^{\mathrm{I}}, x_2^{\mathrm{I}}, \ldots, x_{n_1}^{\mathrm{I}}$ respectiv $y_1^{\mathrm{I}}, y_2^{\mathrm{I}}, \ldots, y_{m_1}^{\mathrm{I}}$ precum și variabilele aleatoare X^{II} și Y^{II} asupra cărora se fac n_2 respectiv m_2 experiențe independente în perioada T_2, rezultatele acestora fiind $x_1^{\mathrm{II}}, x_2^{\mathrm{II}}, \ldots, x_{n_2}^{\mathrm{II}}$ respectiv $y_1^{\mathrm{II}}, y_2^{\mathrm{II}}, \ldots, y_{m_2}^{\mathrm{II}}$.

Se notează

$$\bar{x}_{n_1} = \frac{x_1^{\mathrm{I}} + x_2^{\mathrm{I}} + \ldots + x_{n_1}^{\mathrm{I}}}{n_1}$$

și în mod analog $\bar{x}_{n_2}^{\mathrm{II}}, \bar{y}_{m_1}^{\mathrm{I}}, \bar{y}_{m_2}^{\mathrm{II}}$, precum și

$$i^{(X)} = \frac{\bar{x}_{n_2}}{x_{n_1}} \quad \text{și} \quad i^{(Y)} = \frac{y_{m_2}}{\bar{y}_{m_1}}.$$

Cînd putem spune că diferenţa dintre cei doi indici este datorită factorilor întîmplători? Dacă condiţiile de funcţionare a maşinilor sînt aceleaşi la cele două ateliere, atunci

$$M\left[x_i^{\mathrm{I}}\right] = M\left[y_j^{\mathrm{I}}\right] = a_1;\ M\left[x_i^{\mathrm{II}}\right] = M\left[y_j^{\mathrm{II}}\right] = a_2,$$

$$M\left[x_i^{\mathrm{I}} - a_1\right]^2 = M\left[y_j^{\mathrm{I}} - a_1\right]^2 = \sigma_1^2;\ M\left[x_i^{\mathrm{II}} - a_2\right]^2 = M\left[y_j^{\mathrm{II}} - a_2\right]^2 = \sigma_2^2.$$

În condiţiile problemei, cînd $n_1 \to \infty$, $\bar{x}^{\mathrm{I}}_{n_1}$ tinde către o lege normală $\left[a_1; \frac{\sigma_1}{\sqrt{n_1}}\right]$. În condiţii analoage, $\bar{x}_{n_2}$, $\bar{y}_{n_1}$ şi $\bar{y}_{n_2}$ tind către legi normale:

$$\left[a_2; \frac{\sigma_2}{\sqrt{n_2}}\right], \left[a_1; \frac{\sigma_1}{\sqrt{m_1}}\right], \left[a_2; \frac{\sigma_2}{\sqrt{m_2}}\right].$$

Dacă min $[m_1, n_1, m_2, n_2]$ se ia suficient de mare pentru a se putea aproxima distribuţia exactă cu suficientă precizie prin cea normală, atunci alegînd un t convenabil pentru a asigura un prag de siguranţă dorit, dacă

$$|i^{(X)} - i^{(Y)}| > \max.\left[\left|\frac{a_2 + \frac{t\sigma_2}{\sqrt{n_2}}}{a_1 - \frac{t\sigma_1}{\sqrt{n_1}}} - \frac{a_2 - \frac{t\sigma_2}{\sqrt{m_2}}}{a_1 + \frac{t\sigma_1}{\sqrt{m_1}}}\right|; \left|\frac{a_2 + \frac{t\sigma_2}{\sqrt{m_2}}}{a_1 - \frac{t\sigma_1}{\sqrt{m_1}}} - \frac{a_2 - \frac{t\sigma_2}{\sqrt{n_2}}}{a_1 + \frac{t\sigma_1}{\sqrt{n_1}}}\right|\right]$$

diferenţa dintre cei doi indici este semnificativă.

Desigur a_1, a_2, σ_1, σ_2 sînt de înlocuit prin estimaţiile corespunzătoare ca mai înainte.

Toate acestea sînt valabile numai pentru n_1, n_2, m_1, m_2 suficient de mari pentru ca utilizarea legilor normale limită în locul legilor exacte să fie admisibilă cu o precizie suficientă.

Se stabileşte astfel un criteriu suficient, sub rezerva celor de mai sus, pentru a recunoaşte că diferenţa dintre mărimea a doi indici este semnificativă, deci nu poate fi datorată factorilor aleatori.

Lipsa de bază a criteriilor de mai sus şi în mod deosebit a acelui din urmă este bazarea lor pe legi asimptotice, fără luarea în considerare a efectului erorilor rezultate prin înlocuirea legilor exacte cu cele asimptotice. Cu alte cuvinte nu se vede cît de mari trebuie să fie numerele n_i, m_j pentru ca erorile rezultate prin înlocuirea legilor exacte prin cele asimptotice să fie mici în raport cu diferenţele dintre mărimile celor doi indici, pînă la care acestea pot fi datorate factorilor aleatori.

De aceea vom trece la studiul distribuţiilor exacte şi vom cerceta măsura în care acestea, pentru probe suficient de mari, pot fi aproximate în mod satisfăcător prin legi asimptotice.

Să considerăm două variabile aleatoare X şi Y independente supuse la legi Gauss $(m_1; \sigma_1)$; $(m_2; \sigma_2)$.

Se fac n_1 probe independente relativ la variabila X şi n_2 relativ la Y, rezultatele fiind $x_1, x_2, \ldots, x_{n_1}$ respectiv $y_1, y_2, \ldots, y_{n_2}$.

Se notează

$$\bar{x}_{n_1} = \frac{x_1 + x_2 + \ldots + x_{n_1}}{n_1};\ \bar{y}_{n_2} = \frac{y_1 + y_2 + \ldots + y_{n_2}}{n_2}$$

şi se cere să se afle legea de distribuţie a mărimii aleatoare $I_{X;Y} = \frac{\bar{y}_{n_2}}{\bar{x}_{n_1}}$; $\bar{x}_{n_1}$ şi $\bar{y}_{n_2}$ urmează

legii de distribuție Gauss $\left[m_1 ; \frac{\sigma_1}{\sqrt{n_1}}\right], \left[m_2 ; \frac{\sigma_2}{\sqrt{n_2}}\right]$. Cazul care ne interesează este acel în care $m_1, m_2 > 0$.

Întrucît mai tîrziu vom avea nevoie să stabilim legea de distribuție a unei variabile aleatoare de forma $X = \frac{Y}{Z}$, unde $(Y ; Z)$ urmează legea lui Gauss bidimensională, vom începe direct cu acest caz.

Dacă se consideră o variabilă aleatoare bidimensională $(Y ; Z)$ avînd densitatea de probabilitate $p(y, z)$ și o nouă variabilă aleatoare $X = \frac{Y}{Z}$, atunci:

$$F_X(x) = P\left[\frac{Y}{Z} < x\right] = \int_0^{\infty}\int_{-\infty}^{zx} p(y, z)\, dy\, dz + \int_{-\infty}^{0}\int_{zx}^{\infty} p(y, z)\, dy\, dz.$$

Derivînd în raport cu x, se obține

$$p_X(x) = \int_0^{\infty} z\, p(z\,x, x)\, dz - \int_{-\infty}^{0} z\, p(z\,x ; x)\, dz.$$

În cazul nostru

$$p(y, z) = \frac{1}{2\pi\,\sigma_1\,\sigma_2\sqrt{1-\rho^2}}\, e^{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{y-m_1}{\sigma_1}\right)^2 - 2\rho\frac{(y-m_1)(z-m_2)}{\sigma_1\sigma_2} + \left(\frac{z-m_2}{\sigma_2}\right)^2\right]}.$$

Prin urmare

$$p_X(x) = \frac{1}{2\pi\,\sigma_1\,\sigma_2\sqrt{1-\rho^2}}\left[\int_0^{\infty} z\, e^{-\frac{1}{2(1-\rho_2)}\left[\left(\frac{xz-m_1}{\sigma_1}\right)^2 - 2\rho\left(\frac{xz-m_1}{\sigma_1}\right)\left(\frac{z-m_2}{\sigma_2}\right) + \left(\frac{z-m_2}{\sigma_2}\right)^2\right]} dz.\right.$$

$$\left. - \int_{-\infty}^{0} z\, e^{-\frac{1}{2(1-\rho^1)}\left[\left(\frac{xz-m_1}{\sigma_1}\right)^2 - 2\rho\left(\frac{xz-m_1}{\sigma_1}\right)\left(\frac{z-m_2}{\sigma_2}\right) + \left(\frac{z-m_2}{\sigma_2}\right)^2\right]} dz\right].$$

Expresia de la exponent se mai poate scrie

$$\underbrace{\left[\frac{x^2}{\sigma_1^2} - 2\rho\frac{x}{\sigma_1\,\sigma_2} + \frac{1}{\sigma_2^2}\right]}_{A_1} z^2 - 2\underbrace{\left[\frac{m_1\,x}{\sigma_1^2} - \frac{(m_1 + m_2\,x)\,\rho}{\sigma_1\,\sigma_2} + \frac{m_2}{\sigma_2^2}\right]}_{B_1} z + \underbrace{\left[\frac{m_1^2}{\sigma_1^2} - 2\rho\frac{m_1\,m_2}{\sigma_1\,\sigma_2} + \frac{m_2^2}{\sigma_2^2}\right]}_{C_1}.$$

Este de observat că această formă pătratică este pozitivă pentru toate valorile lui x și z cu excepția cazului $xz - m_1 = 0 ; z - m_2 = 0 ; x = \frac{m_1}{m_2} ; z = m_2$.

Prin urmare $B_1^2 - A_1\,C_1 < 0$ pentru $x \neq \frac{m_1}{m_2}$ iar pentru $x = \frac{m_1}{m_2}$ se anulează.

Notînd: $\frac{1}{2(1-\rho^2)} A_1 = A ; \frac{1}{2(1-\rho^2)} B_1 = B ; \frac{1}{2(1-\rho^2)} C_1 = C$ se deduce:

$$p_X(x) = \frac{1}{2\pi\,\sigma_1\,\sigma_2\sqrt{1-\rho^2}}\left[\int_0^{\infty} z\, e^{-(Az^2 - 2Bz + C)}\, dz - \int_{-\infty}^{0} z\, e^{-(Az^2 - 2Bz + C)}\, dz\right]$$

Cum

$$Az^2 - 2Bz + C = A\left(z - \frac{B}{A}\right)^2 + C - \frac{B^2}{A}$$

dacă se face substituția:

$$\sqrt{A}\left(z - \frac{B}{A}\right) = u\,;\; z = \frac{u}{\sqrt{A}} + \frac{B}{A},$$

paranteza mare devine

$$e^{\frac{B^2}{A}-C}\left[\int_{-\frac{B}{\sqrt{A}}}^{\infty}\left(\frac{u}{\sqrt{A}}+\frac{B}{A}\right)e^{-u^2}\frac{du}{\sqrt{A}} - \int_{-\infty}^{-\frac{B}{\sqrt{A}}}\left(\frac{u}{\sqrt{A}}+\frac{B}{A}\right)e^{-u^2}\frac{du}{\sqrt{A}}\right] =$$

$$= \frac{1}{A}e^{\frac{B^2}{A}-C}\left[-\frac{1}{2}e^{-u^2}\Big|_{-\frac{B}{\sqrt{A}}}^{\infty} - \left(-\frac{1}{2}e^{-u^2}\right)\Big|_{-\infty}^{-\frac{B}{\sqrt{A}}} + \frac{B}{A\sqrt{A}}\left[\int_{-\frac{B}{\sqrt{A}}}^{\infty}e^{-u^2}du - \int_{-\infty}^{-\frac{B}{\sqrt{A}}}e^{-u^2}du\right]\right] =$$

$$= \frac{1}{A}e^{\frac{B^2}{A}-C}\left[e^{-\frac{B^2}{A}} + \frac{B}{\sqrt{A}}\int_{-\frac{B}{\sqrt{A}}}^{\frac{B}{\sqrt{A}}}e^{-u^2}du\right] = \frac{1}{A}e^{-C}\left[1 + \frac{B^2}{A}\int_{-1}^{1}e^{\frac{B^2}{A}(1-t^2)}dt\right].$$

Ținînd seama de faptul că

$$A = \frac{1}{2(1-\rho^2)}\left[\frac{x^2}{\sigma_1^2} - 2\rho\frac{x}{\sigma_1\sigma_2} + \frac{1}{\sigma_2^2}\right]$$

se deduce imediat

$$p_X(x) = \frac{\sqrt{1-\rho^2}\,\sigma_1\sigma_2}{\pi\left[\sigma_2^2x^2 - 2\rho\sigma_1\sigma_2x + \sigma_1^2\right]}\,e^{-\frac{1}{2(1-\rho^2)}\left[\frac{m_1^2}{\sigma_1^2} - 2\rho\frac{m_1m_2}{\sigma_1\sigma_2} + \frac{m_2^2}{\sigma_2^2}\right]}\cdot\left[1 + 2\frac{B^2}{A}\int_0^1 e^{\frac{B^2}{A}(1-t^2)}dt\right],$$

unde

$$B = \frac{1}{2(1-\rho^2)}\left|\left(\frac{m_1}{\sigma_1^2} - \frac{m_2\rho}{\sigma_1\sigma_2}\right)x + \left(\frac{m_2}{\sigma_1^2} - \frac{m_1\rho}{\sigma_1\sigma_2}\right)\right|$$

iar expresia lui A a fost dată mai sus.

În cazul particular în care $m_1 = m_2 = 0$ deci în care $B = C = 0$; $A \neq 0$ se obține rezultatul cunoscut:

$$p_X(x) = \frac{\sqrt{1-\rho^2}\sigma_1\sigma_2}{\pi\left[\sigma_2^2x^2 - 2\rho\sigma_1\sigma_2x + \sigma_1^2\right]},$$

adică distribuția lui Cauchy.

În cazul general în care m_1, m_2, σ_1, σ_2 sînt numere determinate iar m_1, m_2 nu sînt nuli simultan, distribuția deși nu mai este o distribuție Cauchy are însă caracteristici apropiate de ale acesteia.

Într-adevăr, mai înainte s-a arătat că $B_1^2 - A_1C_1 \leqslant 0$, $\ldots\ \frac{B^2}{A} \leqslant C$, unde C este un număr pozitiv determinat:

$$C = \frac{1}{2(1-\rho^2)}\left[\frac{m_1^2}{\sigma_1^2} - 2\rho\frac{m_1m_2}{\sigma_1\sigma_2} + \frac{m_2^2}{\sigma_2^2}\right].$$

Prin urmare,

$$2 < \int_0^1 e^{\frac{B^2(x)}{A(x)}[1-t^2]}\,dt < \underbrace{\int_0^1 e^{C(1-t^2)}\,dt}$$

număr pozitiv determinat.

Aceasta înseamnă că notînd:

$$2\frac{B^2(x)}{A(x)}\int_0^1 e^{\frac{B^2(x)}{A(x)}[1-t^2]}\,dt = \Phi(x),$$

$\Phi(x)$ este o funcție definită pentru toate valorile lui x și mărginită.

Se vede deci că

$$p_X(x) = \frac{\sqrt{1-\rho^2}\,\sigma_2\sigma_2}{\pi[\sigma_2^2x^2 - 2\rho\sigma_1\sigma_2x + \sigma_1^2]}\,e^{-C}\,[1+\Phi(x)].$$

Structura acestei densități a funcției de distribuție este de același tip ca a distribuției Cauchy, prezentînd același mare dezavantaj de a nu avea momente întregi pozitive. Nu se poate vorbi nici de valoare probabilă (moment de ordinul 1) nici de dispersie.

Problemele de estimație nu pot fi abordate aici după metodele obișnuite.

Există însă momente de ordin fracționar pozitiv.

Dacă s-ar estima în acest mod, destul de incomod pentru practică, un moment de ordin fracționar ar fi necesar să se calculeze apoi și dispersia acestuia. De aceea exponentul fracționar k ar trebui să fie astfel ca $2k < 1$; $k < \frac{1}{2}$. De exemplu s-ar putea calcula valoarea medie a lui $\sqrt[3]{X}$ și dispersia acestuia, deci,

$$M\left[\sqrt[3]{X}\right] \text{ și } M\left[\sqrt[3]{X} - M\left(\sqrt[3]{X}\right)\right]^2.$$

În problema care ne preocupă, avem de-a face cu o altă situație decît cea de mai sus, care se exprimă prin faptul că m_1 și m_2 sînt numere determinate dar abaterile pătratice medii nu mai apar ca numere constante, ci ca mărimi variabile, care tind la zero cînd numărul probelor tinde la infinit.

Astfel, cînd am vorbit despre indicele $I_{X:Y} = \frac{\bar{y}_{n_2}}{\bar{x}_{n_1}}$ am văzut că $\bar{y}_{n_2}$ era supus unei legi normale $\left[m_2; \frac{\sigma_2}{\sqrt{n_2}}\right]$ iar $\bar{x}_{n_1}$ unei legi normale $\left[m_1; \frac{\sigma_1}{\sqrt{n_1}}\right]$.

Dar lucrurile nu se prezintă astfel numai cînd fiecare componentă este supusă legii normale.

Teorema lui Liapunov corespunzătoare cazului bidimensional prezintă un caz mult mai important.

Dacă se urmărește activitatea unui aceluiași grup de persoane la două etape diferite și se consideră activitatea diferitelor persoane în primă aproximație ca realizări independente ale unei aceleiași variabile aleatoare X, nu același lucru se poate spune despre activitatea unei aceleiași persoane în două etape diferite.

Vom avea de urmărit diferitele realizări ale variabilei bidimensionale $(X; Y)$ și anume $(x_1; y_1)$, $(x_2; y_2)$,..., $(x_n; y_n)$ presupunîndu-se probele independente între ele. Dacă se notează:

$$M[x_i] = m_1;\ M[y_j] = m_2;\ M[x_i - m_1]^2 = \sigma_1^2;\ M[y_j - m_2]^2 = \sigma_2^2:$$

$$M[(x_i - m_1)(y_j - m_2)] = \rho\,\sigma_1\sigma_2,$$

atunci deşi mărimile aleatoare $(x_i\,;\,y_i)$ nu sînt supuse legii normale, întrucît însă în cazul problemei noastre

$$\max_{i=1;2;\ldots n}\left[\,|\,x_i - m_1\,|\,;\,|\,y_i - m_2\,|\,\right] < A,$$

unde A este un număr determinat care nu depinde de n, distribuţia mărimii aleatoare $\left[\frac{x_1 + x_2 + \ldots + x_n}{n}\,;\,\frac{y_1 + y_2 + \ldots + y_n}{n}\right]$ tinde, cînd $n \to \infty$, către legea lui Gauss bidimensională $\left[m_1\,;\,m_2\,;\,\frac{\sigma_1}{\sqrt{n}}\,;\,\frac{\sigma_2}{\sqrt{n}}\,;\,\rho\right]$.

Să considerăm acum cazul în care în expresia lui $p_X(x)$, m_1, m_2 sînt numere fixe ca şi ρ, dar în loc de σ_1, σ_2 se ia $\frac{\sigma_1}{\sqrt{n}}$, $\frac{\sigma_2}{\sqrt{n}}$.

Se vede imediat că în acest caz

$$p(x) = \frac{\sqrt{1-\rho^2}\,\sigma_1\sigma_2}{\pi\left[\sigma_2^2x^2 - 2\rho\sigma_1\sigma_2x + \sigma_1^2\right]}\,e^{-\frac{n}{2(1-\rho^2)}\left[\frac{m_1^2}{\sigma_1^2} - 2\rho\frac{m_1m_2}{\sigma_1\sigma_2} + \frac{m_2^2}{\sigma_2^2}\right]}\left[1 + \frac{2n\,B^2}{A}\int_0^1 e^{n\frac{B^2}{A}(1-t^2)}\,dt\right].$$

Ne propunem acum să calculăm primii termeni ai funcţiei de repartiţie corespunzătoare:

$$F(x) = \int_{-\infty}^{x} p(x)\,dx$$

adică partea principală şi primii termeni după puterile eventual fracţionare ale lui $\frac{1}{n}$.

În acest scop observînd că

$$\int_{-\infty}^{\infty} \frac{\sqrt{1-\rho^2}\,\sigma_1\sigma^2}{\pi\left[\sigma_2^2x^2 - 2\rho\sigma_1\sigma_2x + \sigma_1^2\right]}\,dx = 1\,;$$

$p(x)$ mai poate fi scris

$$p(x) = \frac{\sqrt{1-\rho^2}\,\sigma_1\sigma_2}{\pi\left[\sigma_2^2x^2 - 2\rho\sigma_1\sigma_2x + \sigma_1^2\right]}\,e^{-nC} + \frac{2\sqrt{1-\rho^2}\,\sigma_1\sigma_2\,nB^2}{\pi\left[\sigma_2^2x^2 - 2\rho\sigma_1\sigma_2x + \sigma_1^2\right]A}\,e^{n\left[\frac{B^2(x)}{A(x)} - C\right]}\int_0^1 e^{-n\frac{B^2(x)}{A(x)}t^2}\,dt.$$

Însă cînd se calculează $F(x)$, contribuţia primului termen din membrul din dreapta egalităţii e mai mică decît e^{-nC} deci mai mică decît orice putere a lui $\frac{1}{n}$ astfel că nu este de luat în cazul nostru în considerare.

În ceea ce priveşte al doilea termen, întrucît $\frac{B^2(x)}{A(x)} \leqslant C$ oricare ar fi x iar

$$\int_0^1 e^{-n\frac{B^2(x)}{A(x)}t^2}\,dt < 1,$$

acesta aduce o contribuţie în limitele propuse numai în măsura în care $n\left[\frac{B^2(x)}{A(x)} - C\right]$ nu tinde la $-\infty$ cînd $n \to \infty$.

Prin urmare, va trebui să ne concentrăm atenţia numai asupra valorilor $x = x(n)$ pentru care expresia $n\left[\frac{B^2(x)}{A(x)} - C\right]$ tinde cel mult către o limită finită cînd $n \to \infty$.

E de prevăzut totodată că pentru a se obţine o probabilitate oricît de apropiată de unitate va fi suficient să calculăm:

$$\int_{x_1(n)}^{x_2(n)} p(x)\, \mathrm{d}\, x$$

cu $x_1(n)$ şi $x_2(n)$ aleşi în mod convenabil.

Însă am văzut că $\frac{B^2(x)}{A(x)} - C = 0$ pentru $x = \frac{m_1}{m_2}$, de aceea vom face substituţia

$$x(n) = \frac{m_1}{m_2} + y\,(n)$$

şi vom determina valorile maxime şi minime ale lui $y\ (n)$ astfel ca

$$\lim_{n \to \infty} n\left[\frac{B^2\left(\frac{m_1}{m_2} + y\right)}{A\left(\frac{m_1}{m_2} + y\right)} - C\right] = -a,$$

a fiind un număr pozitiv dat.

Însă

$$A_1 = \frac{1}{m_2^2}\left[C_1 + \frac{2m_2 y}{\sigma_1}\left(\frac{m_1}{\sigma_1} - \frac{\rho m_2}{\sigma_2}\right) + \frac{m_2^2 y^2}{\sigma_1^2}\right],$$

$$B_1 = \frac{1}{m_2}\left[C_1 + \frac{m_2 y}{\sigma_1}\left(\frac{m_1}{\sigma_1} - \frac{\rho m_2}{\sigma_2}\right)\right],$$

unde după cum ştim

$$C_1 = \left[\frac{m_1^2}{\sigma_1^2} - 2\rho\,\frac{m_1 m_2}{\sigma_1\sigma_2} + \frac{m_2^2}{\sigma_2^2}\right].$$

Efectuînd calculele se obţine

$$\frac{B_1^2}{A_1} - C_1 = -\frac{(1-\rho^2)\, m_2^4\, y^2}{\sigma_1^2\sigma_2^2} \cdot \frac{1}{C_1 + \frac{2m_2 y}{\sigma_1}\left(\frac{m_1}{\sigma_1} - \rho\frac{m_2}{\sigma_2}\right) + \frac{m_2^2 y^2}{\sigma_2^2}}.$$

E de observat că

$$\frac{B^2}{A} - C = \frac{1}{2(1-\rho^2)}\left[\frac{B_1^2}{A_1} - C_1\right].$$

Prin urmare, pentru ca

$$\lim_{n \to \infty} n\left[\frac{B^2\left(\frac{m_1}{m_2} + y\right)}{A\left[\frac{m_1}{m_2} + y\right]} - C\right]$$

să existe şi să fie un număr finit trebuie ca $\lim\limits_{n \to \infty} ny^2\,(n)$ să existe şi să fie un număr finit.

Vedem astfel că pentru valorile extreme ale lui y s-ar putea lua numere de forma $-\frac{k}{\sqrt{n}}$ și $\frac{k}{\sqrt{n}}$, unde k e un număr determinat. Desigur e de prevăzut că numărul k va putea fi ales astfel încît probabilitatea inegalității $-\frac{k}{\sqrt{n}} < x - \frac{m_1}{m_2} < \frac{k}{\sqrt{n}}$ să fie oricît de apropiată de unitate.

Sîntem astfel conduși a calcula primii termeni ai dezvoltării integralei

$$\int_{\frac{m_1}{m_2}-\frac{k}{\sqrt{n}}}^{\frac{m_1}{m_2}+\frac{k}{\sqrt{n}}} p^{II}(x)\,dx = \int_{-\frac{k}{\sqrt{n}}}^{\frac{k}{\sqrt{n}}} p^{II}(x)\left(\frac{m_1}{m_2}+y\right)dy$$

după puterile crescătoare ale lui $\frac{1}{\sqrt{n}}$. Aici $p^{II}(x)$ reprezintă al doilea termen al descompunerii lui $p(x)$, care singur e de considerat, după cum s-a arătat mai înainte.

Însă:

$$p^{II}\left(\frac{m_1}{m_2}+y\right) = \frac{2\sqrt{1-\rho^2}\,\sigma_1\sigma_2 n \frac{1}{4(1-\rho^2)^2} B_1^2\left(\frac{m_1}{n_1}+y\right)}{\pi\left[\sigma_2^2\left(\frac{m_1}{m_2}+y\right)^2 - 2\rho\,\sigma_1\sigma_2\left(\frac{m_1}{m_2}+y\right)+\sigma_1^2\right]\cdot\frac{1}{2(1-\rho^2)}A_1\left(\frac{m_1}{n_1}+y\right)} \times$$

$$\times\, e^{\frac{n}{2(1-\rho^2)}\left[\frac{B_1^2}{A_1}-C_1\right]}\int_0^1 e^{-\frac{n}{2(1-\rho^2)}\frac{B_1^2}{A_1}t^2}\,dt =$$

$$= \frac{\sigma_1\sigma_2 n B_1^2}{\pi\left[\sigma_2^2\left(\frac{m_1}{m_2}+y\right)^2 - 2\rho\,\sigma_1\sigma_2\left(\frac{m_1}{m_2}+y\right)+\sigma_1^2\right]A_1\sqrt{1-\rho^2}} \times$$

$$\times\, e^{\frac{n}{2(1-\rho^2)}\left[\frac{B_1^2}{A_1}-C_1\right]}\int_0^1 e^{-\frac{n}{2(1-\rho^2)}\frac{B_1^2}{A_1}t^2}\,dt.$$

Ne vom mărgini la aflarea termenului principal. Aflarea altor termeni se reduce la o problemă de calcul, neîntroducîndu-se nimic principial nou.

Întrucît

$$\lim_{n\to\infty} n\left|\frac{B^2}{A}-C\right| = \text{a număr finit}$$

rezultă că

$$\frac{B_1^2}{A_1} \to C_1$$

sau mai precis

$$\frac{B_1^2}{A_1} = C_1 + 0\left(\frac{1}{n}\right).$$

Dar atunci $\frac{B_1^2}{A_1}$ poate fi înlocuit ținînd seama de precizia urmărită cu C_1.

Apoi:

$$\int_0^1 e^{-\frac{n}{2(1-\rho^2)}\frac{B_1^2}{A_1}t^2}\,dt \simeq \int_0^1 e^{\frac{n}{2(1-\rho^2)}C_1 t^2}\,dt =$$

$$= \frac{\sqrt{1-\rho^2}}{\sqrt{n}\sqrt{C_1}}\int_0^{\sqrt{\frac{n}{1-\rho^2}}\sqrt{C_1}} e^{-\frac{t^2}{2}}\,dt \cong \sqrt{\frac{\pi}{2}}\frac{\sqrt{1-\rho^2}}{\sqrt{n}\sqrt{C_1}} + 0\left(\frac{1}{n^p}\right) \text{ cu } p \text{ oricît de mare.}$$

Apoi cum

$$\frac{B_1^2}{A_1} - C_1 = -\frac{(1-\rho^2)}{\sigma_1^2\sigma_2^2}\frac{m_2^4 y^2}{C_1 + \frac{2m_2 y}{\sigma_1}\left(\frac{m_1}{\sigma_1} - \rho\frac{m_2}{\sigma_2}\right) + \frac{m_2^2 y^2}{\sigma_2^2}}$$

se vede că partea principală este

$$-\frac{(1-\rho^2)m_2^4 y^2}{\sigma_1^2\sigma_2^2 C_1},$$

abaterea de la valoarea exactă fiind de ordinul $\frac{1}{\sqrt{n}}$.

Partea principală a integralei este deci:

$$\int_{-\frac{k}{\sqrt{n}}}^{\frac{k}{\sqrt{n}}} \frac{\sigma_1\sigma_2 n C_1}{\pi\left|\sigma_2^2\frac{m_1^2}{m_2^2} - 2\rho\sigma_1\sigma_2\frac{m_1}{m_2} + \sigma_1^2\right|\sqrt{1-\rho^2}} \cdot \sqrt{\frac{\pi}{2}}\cdot\frac{\sqrt{1-\rho^2}}{\sqrt{n}\sqrt{C_1}} e^{-\frac{m_2^4 y^2 n}{2\sigma_1^2\sigma_2^2 C_1}}\,dy =$$

$$= \frac{\sigma_1\sigma_2 C_1 n}{\sqrt{2\pi}\frac{\sigma_1^2\sigma_2^2}{m_2^2}C_1\sqrt{C_1}}\int_{-\frac{k}{\sqrt{n}}}^{\frac{k}{\sqrt{n}}} e^{-\frac{m_2^4 y^2 n}{2\sigma_1^2\sigma_2^2 C_1}}\,dy.$$

Făcînd schimbarea de variabilă

$$\frac{m_2^2\sqrt{n}}{\sigma_1\sigma_2\sqrt{C_1}}y = z$$

și înlocuind pe k prin

$$\frac{\sigma_1\sigma_2\sqrt{C_1}}{m_2^2}a$$

se obține pentru probabilitate expresia:

$$\frac{1}{\sqrt{2\pi}}\int_{-a}^{a} e^{-\frac{z^2}{2}}\,dz,$$

adică distribuția limită normală.

Am obținut astfel următorul rezultat:
probabilitatea inegalității

$$\left|\frac{Y}{z} - \frac{m_1}{m_2}\right| < a\frac{\sigma_1\sigma_2\sqrt{C_1}}{m_2^2\sqrt{n}},$$

unde

$$C_1 = \frac{m_1^2}{\sigma_1^2} - 2\rho \frac{m_1 m_2}{\sigma_1 \sigma_2} + \frac{m_2^2}{\sigma_2^2},$$

este

$$\frac{1}{\sqrt{2\pi}} \int_{-a}^{a} e^{-\frac{z^2}{2}} dz + 0\left(n^{-\frac{1}{2}}\right).$$

În particular, cu o probabilitate egală cu 0,9973 a fost stabilită deci relația

$$\left| \frac{Y}{Z} - \frac{m_1}{m_2} \right| < \frac{3\,\sigma_1\sigma_2\sqrt{C_1}}{m_2^2\sqrt{n}}.$$

În cazul particular în care $m_1 = 0$,

$$\frac{\sigma_1 \sigma_2 \sqrt{C_1}}{m_2^2 \sqrt{n}} = \frac{\sigma_1}{m_2 \sqrt{n}}$$

și se regăsește un rezultat obținut pe altă cale mai înainte.

Dacă ne limităm la expresii asimptotice și la cazul în care variabilele aleatoare X și Y sînt independente, atunci fără a presupune că X și Y sînt supuse legii normale, se obțin rezultate de seamă pentru indicele $\frac{\overline{y}_n}{\overline{x}_{n_1}}$, dacă $\frac{n_2}{n_1+n_2} \to k$, număr finit, cînd $n_1, n_2 \to \infty$.

Într-adevăr, să presupunem că

$$M(x_i) = m_1;\ M(x_i - m_1)^2 = \sigma_1^2;\ M[y_j] = m_2;\ M(y_j - m_2)^2 = \sigma_2^2.$$

Atunci

$$\sqrt{\frac{n_1 n_2}{n_1 + n_2}}\left[\frac{\overline{y}_{n_2}}{\overline{x}_{n_1}} - \frac{m_2}{m_1}\right] = \frac{\displaystyle\sum_{j=1}^{n_2} \sqrt{\frac{n_1 n_2}{n_1+n_2}}\left[\frac{m_1(y_j - m_2)}{n_2}\right] + \sum_{i=1}^{n_1} \sqrt{\frac{n_1 n_2}{n_1+n_2}}\left[-\frac{m_2(x_i - m_1)}{n_1}\right]}{\displaystyle\sum_{i=1}^{n_1} \frac{m_1 x_i}{n_1}}.$$

La numărător avem o sumă de $(n_1 + n_2)$ mărimi independente pentru care

$$M\left[\sqrt{\frac{n_1 n_2}{n_1+n_2}}\,\frac{m_1(y_j - m_2)}{n_2}\right] = 0,\ M\left[\sqrt{\frac{n_1 n_2}{n_1+n_2}}\,\frac{m_1(y_j - m_2)}{n_2}\right]^2 = \frac{n_1}{n_2(n_1+n_2)} m_1^2 \sigma_2^2,$$

$$M\left[\sqrt{\frac{n_1 n_2}{n_1+n_2}}\,\frac{m_2(x_i - m_1)}{n_1}\right] = 0;\ M\left[\sqrt{\frac{n_1 n_2}{n_1+n_2}}\,\frac{m_1(x_i - m_1)}{n_1}\right]^2 = \frac{n_2}{n_1(n_1+n_2)} m_2^2 \sigma_1^2.$$

Suma dispersiilor la numărător este

$$n_2 \frac{n_1}{n_2(n_1+n_2)} m_1^2 \sigma_2^2 + n_1 \frac{n_2}{n_1(n_1+n_2)} m_2^2 \sigma_1^2 = \frac{n_2}{n_1+n_2} m_2^2 \sigma_1^2 + \frac{n_1}{n_1+n_2} m_1^2 \sigma_2^2.$$

Făcînd $\frac{n_1 n_2}{n_1+n_2} \to \infty$, dispersia tinde către $k\, m_2^2 \sigma_1^2 + (1-k)\, m_1^2 \sigma_2^2$.

Pe baza teoremei lui Liapunov ale cărei condiții de aplicare sînt evident satisfăcute în cazul nostru, mărimea variabilă de la numărătorul fracției de mai sus este asimptotic normală $\left[0;\ \sqrt{k\, m_2^2 \sigma_1^2 + (1-k)\, m_1^2 \sigma_2^2}\right]$.

Scanned with OKEN Scanner

Mărimea aleatoare de la numitor are media $m_1 \sum_{i=1}^{n_1} M\left(\frac{x_i}{n_1}\right) = m_1^2$ și dispersia $m_1^2 \sum_{i=1}^{n_1} M\left[\frac{x_1 - m_1}{n_1}\right]^2 = \frac{m_1^4}{n_1}$. Ea este asimptotic normală pe baza teoremei lui Liapunov $\left[m_1^2; \frac{m_1^2}{\sqrt{n_1}}\right]$ și tinde deci în probabilitate către m_1^2.

Mărimea variabilă $\sqrt{\frac{n_1 n_2}{n_1 + n_2}}\left[\frac{\bar{y}_{n_2}}{\bar{x}_{n_1}} - \frac{m_2}{m_1}\right]$ este asimptotic normală

$$\left[0; \frac{\sqrt{km_2^2 \sigma_1^2 + (1-k) m_1^2 \sigma_2^2}}{m_1^2}\right].$$

Pentru $\frac{n_2}{n_1 + n_2} \to k$ cînd $n_1, n_2 \to \infty$, probabilitatea inegalității

$$\left| \frac{\bar{y}_{n_2}}{\bar{x}_{n_1}} - \frac{m_2}{m_1} \right| < t \sqrt{\frac{n_1 + n_2}{n_1 n_2}} \cdot \frac{\sqrt{km_2^2 \sigma_1^2 + (1-k) m_1^2 \sigma_2^2}}{m_1^2}$$

este

$$\frac{1}{\sqrt{2\pi}} \int_{-t}^{t} e^{-\frac{x^2}{2}} dx.$$

De exemplu pentru $t = 3$, ea este 0,9973.

Dar nu numai atît.

Sîntem în situația de a putea stabili limitele de variație a diferenței a doi indici $\frac{\bar{y}_{n_2}}{\bar{x}_{n_1}}$ și $\frac{\bar{y}'_{n_2}}{\bar{x}'_{n_1}}$ corespunzători la două experiențe diferite, independente între ele și aceasta cu orice prag de siguranță dorit.

Pentru aceasta observăm că

$$\sqrt{\frac{n_1 n_2}{n_1 + n_2}}\left[\frac{\bar{y}_{n_2}}{\bar{x}_{n_1}} - \frac{\bar{y}'_{n_2}}{\bar{x}'_{n_1}}\right] = \underbrace{\sqrt{\frac{n_1 n_2'}{n_1 + n_2}}\left[\frac{\bar{y}_{n_2}}{\bar{x}_{n_1}} - \frac{m_2}{m_1}\right]}_{X_n^{(1)}} - \underbrace{\sqrt{\frac{n_1 n_2}{n_1 + n_2}}\left[\frac{\bar{y}'_{n_2}}{\bar{x}'_{n_1}} - \frac{m_2}{m_1}\right]}_{X_n^{(2)}}.$$

Dacă se notează funcțiile caracteristice ale lui $X_n^{(1)}$ și $X_n^{(2)}$ cu $\varphi_n(t)$ și $\varphi'_n(t)$, funcția caracteristică a diferenței $X_n^{(1)} - X_n^{(2)}$ va fi $\varphi_n(t)\,\varphi'_n(-t)$, întrucît $X_n^{(1)}$ și $X_n^{(2)}$ sînt independente.

Deoarece $X_n^{(1)}$ și $X_n^{(2)}$ sînt asimptotic normale cu media 0, rezultă

$$\lim_{n\to\infty} \varphi_n(t) = \lim_{n\to\infty} \varphi_n(-t) = \sigma(t),$$

$\sigma(t)$ fiind funcția caracteristică a legii normale

$$\left[0, \frac{\sqrt{km_2^2 \sigma_1^2 + (1-k) m_1^2 \sigma_2^2}}{m_1^2} = \Sigma\right].$$

Cu alte cuvinte, $X_n^{(1)} - X_n^{(2)}$ este asimptotic normală $[0, 2\Sigma]$.

Scanned with OKEN Scanner

Rezultă deci că dacă $\frac{n_2}{n_1+n_2} \to k$ cînd $n_1, n_2 \to \infty$, probabilitatea inegalității:

$$\left| \frac{\bar{y}_{n_2}}{\bar{x}_{n_1}} - \frac{\bar{y}'_{n_2}}{\bar{x}'_{n_1}} \right| < t \sqrt{\frac{n_1+n_2}{n_1 n_2}} \cdot \frac{\sqrt{2\,[km_2^2\sigma_1^2 + (1-k)\,m_1^2\sigma_2^2}}{m_1^2}$$

este

$$\frac{1}{\sqrt{2\pi}} \int_z^t e^{-\frac{x^2}{2}}\,dx.$$

Desigur σ_1^2, σ_2^2, m_1 nu sînt cunoscute. Dar după cum s-a menționat și mai înainte se cunosc estimații nedeplasate cu dispersia minimă pentru acestea, care pot fi utilizate cu suficientă siguranță.

O analiză mai aprofundată ar cere estimația globală a mărimei aleatoare

$$\frac{\sqrt{km_2^2\sigma_1^2 + (1-k)m_1^2\sigma_2^2}}{m_1^2}$$

pe care nu o vom face aici.

Tipul de raționament de mai sus arată că trecerea de la o variabilă aleatoare X la un număr determinat s de variabile $X_1, X_2, \ldots, X_s$ și corespunzător s variabila aleatoare $Y_1, Y_2, \ldots, Y_s$ și la probele corespunzătoare de volum $n_1^1, n_2^1, \ldots, n_s^1$; $n_1^2, n_2^2, \ldots, n_s^2$ nu prezintă nici o dificultate.

Un caz interesant este acel în care

$$m_1 = m_2 = m\ ;\ n_1 = n_2 = n\ ;\ \sigma_1 = \sigma_2 = \alpha m.$$

În acest caz cu o probabilitate de 0,9973 au loc relațiile asimptotice:

$$\left| \frac{\bar{y}_n}{\bar{x}_n} - 1 \right| < \frac{3\sqrt{2}\,\alpha}{\sqrt{n}}; \quad \left| \frac{\bar{y}_n}{\bar{x}_n} - \frac{\bar{y}'_n}{\bar{x}'_n} \right| < \frac{6\alpha}{\sqrt{n}}.$$

De exemplu dacă am considera un grup de 4 900 mașini care funcționează în aceleași condiții și am presupune că în două perioade condițiile nu s-au schimbat iar $\sigma = 0{,}4\ m$, atunci ar fi perfect plauzibil ca indicii să nu fie exact 1, ci să satisfacă relațiile:

$$\left| \frac{\bar{y}_{4900}}{\bar{x}_{4900}} - 1 \right| < \frac{3\sqrt{2}\cdot 0{,}4}{\sqrt{4\,900}} = 2{,}4\,\%\ ; \left| \frac{\bar{y}_{4900}}{\bar{x}_{4900}} - \frac{\bar{y}'_{4900}}{\bar{x}'_{4900}} \right| < 3{,}4\,\%.$$

Aceasta înseamnă că indicile poate să fie cuprins între 0,976 și 1, 024 și că doi coeficienți luați la două întreprinderi pot diferi cu 3,4 %, fără ca să avem diferențe semnificative.

Scanned with OKEN Scanner

PARTEA A CINCEA

METODA SELECŢIEI

Scanned with OKEN Scanner

Scanned with OKEN Scanner

Capitolul XIX

Metoda selecţiei şi principiul ei. Importanţa aplicării ei în cercetările statistice

1. Consideraţii generale. Selecţia şi principiul ei

Am văzut că o colectivitate statistică este caracterizată, din punct de vedere statistic, prin distribuţia ei, prin aceasta înţelegîndu-se repartiţia unităţilor sau elementelor colectivităţii după mărimea concretă a unei caracteristici luate în studiu. De pildă, colectivitatea populaţiei unei ţări poate fi distribuită pe grupe de vîrstă, pe sexe, pe medii — urban şi rural — etc. Putem avea însă şi mai multe caracteristici după care se distribuie colectivităţile statistice.

În capitolul IX am dat numeroase exemple de colectivităţi statistice şi distribuţiile lor. Tot atunci am arătat cele două mari categorii în care pot fi împărţite colectivităţile statistice : colectivităţi finite şi colectivităţi infinite şi că în viaţa social-economică ne întîlnim, de obicei, cu colectivităţi din prima categorie, caracterizate prin faptul că numărul unităţilor care le compun este finit.

Pentru colectivităţi finite este posibilă o înregistrare totală şi apoi prelucrarea tuturor unităţilor, în vederea determinării anumitor date privind caracteristicile luate în studiu. Acest lucru nu mai este cu putinţă, din punct de vedere practic, pentru a doua categorie de colectivităţi. Prelucrarea totală a datelor privind aceste colectivităţi nemaifiind posibilă, s-a recurs în mod firesc la o înregistrare şi prelucrare parţială, *extrăgîndu-se, la întîmplare, o subcolectivitate din colectivitatea considerată* şi aplicînd metodele statistice de prelucrare a datelor la această subcolectivitate. Acesta este principiul selecţiei. El se bazează în întregime pe faptul că colectivitatea la care se aplică metodele statistice de prelucrare a datelor, este „aleasă la întîmplare“ din colectivitatea totală şi că orice altă colectivitate cu acelaşi număr de unităţi, are aceeaşi şansă să se ivească

în aplicarea practică a acestei metode. Aceasta revine la presupunerea că fiecare unitate are aceeași șansă de a intra în subcolectivitatea de selecție. În limbajul teoriei probabilităților putem spune că toate unitățile au aceeași probabilitate de a face parte din subcolectivitatea de selecție.

Pe baza datelor obținute prin aplicarea metodei selecției *) se stabilesc apoi indicatorii statistici necesari pentru a caracteriza întreaga colectivitate considerată.

Este evident că metoda selecției poate fi aplicată cu succes și în studiul colectivităților care, practic, pot fi considerate ca avînd un număr infinit de unități. Astfel, colectivitatea formată din globulele de diferite categorii din sîngele existent în corpul unui om, colectivitățile formate din particulele elementare ale corpurilor din fizica statistică sau colectivități cuprinzînd rezervele subterane de minereuri, țiței etc. rezervele de material lemnos de diferite esențe ale fondului silvic al țării etc. nu se pot stabili prin observarea totală. De asemenea, metoda selecției poate fi aplicată în toate împrejurările în care înregistrarea totală nu se poate utiliza fără prejudiciul ireparabil al obiectelor studiate (cercetarea calității diferitelor produse : măsurarea duratei medii de ardere a becurilor, încercarea rezistenței la rupere a firelor etc.) sau în care înregistrarea totală necesită un volum considerabil de cheltuieli de timp de muncă și de materiale ; în general, ea este aplicabilă în toate cazurile cînd prezintă avantaje mai mari decît o cercetare totală, oferind posibilitatea utilizării unui program mult mai bogat, mai complex și — adeseori — chiar un grad de exactitate mai mare a rezultatelor obținute. Deși pare paradoxal, totuși lucrul acesta este cu putință, deoarece folosirea unui număr mai mic de date statistice permite mai ușor eliminarea diverselor erori, care se pot întîlni atît în culegerea datelor, cît și în prelucrarea lor.

Larga aplicare a metodei selecției se explică prin aspectul ei intuitiv care face ca persoane, care de multe ori nu cunosc teoria statisticii, să aibă încredere în rezultatele ei. Astfel, în medicină, putem considera că toate analizele medicale se bazează pe metoda selecției. Majoritatea cercetărilor statistice medicale se referă la o subcolectivitate selectivă de bolnavi. Rezultatele obținute se consideră, prin extindere, că sînt valabile pentru colectivitatea întreagă de bolnavi, fără ca cercetătorii respectivi să se preocupe prea mult de regulile statisticii.

De asemenea, este intuitiv faptul că rezultatele obținute prin metoda selecției sînt cu atît mai precise cu cît numărul unităților din subcolectivitatea de selecție este mai mare.

Întuiția aduce servicii considerabile în știință. Însă, folosirea procedeelor intuitive, fără un control științific, poate conduce la erori grave. De aceea, aplicarea metodei selecției trebuie să fie făcută cu multă grijă, iar rezultatele ei admise numai după o prealabilă analiză științifică.

O interesantă problemă se pune în legătură cu perspectiva dezvoltării continue a teoriei și practicii metodei selecției.

*) În literatura de specialitate se mai folosesc și expresiile „metoda eșantioanelor", „metoda sondajelor" sau simplu „sondaj" și „metoda selectivă". În această lucrare vom folosi denumirea din text sau, pe scurt, „sondaj" cînd ne vom referi la subcolectivități de selecție.

Scanned with OKEN Scanner

Se știe că, în ultimii ani, știința și tehnica au făcut progrese imense.

Oamenii de știință au reușit să construiască mașini electronice de calcul care efectuează zeci de mii de operații pe secundă. Într-un articol publicat de acad. A. Nesmeianov, cu privire la sarcinile științei sovietice în planul septenal *), se afirmă că pînă în prezent s-au construit mașini de calcul care efectuează zeci de mii de operații pe secundă, dar că știința și tehnica vor avea nevoie în curînd de mașini care să facă sute de mii sau poate un milion de operații pe secundă. Această mare putere de calcul, care merge progresînd, va influența oare asupra teoriei și practicii statisticii?

Această întrebare se pune în mod natural dacă ne referim în special la teoria selecției, al cărei avantaj constă, printre altele, în economisirea timpului destinat calculelor, fapt care în curînd s-ar putea să nu mai aibă mare importanță.

Este indiscutabil că, în practica statisticii, mașinile de calcul vor introduce o rapiditate în execuție și o precizie în rezultate cu totul remarcabile. Statistica se bazează în primul rînd pe numărarea cazurilor care compun frecvențele diferitelor fenomene. Operațiile de numărare a cazurilor, de adunare a lor, de determinare a valorilor relative etc., care intervin în mod curent la întocmirea tabelelor statistice, sînt în general simple și se pretează foarte bine de a fi automatizate.

Mașinile electronice fiind în măsură să execute însă rapid și calcule mult mai complicate, încă de pe acum au apărut și alte aplicații ale calculelor matematice în economie, dintre care unele au și început să se constituie în discipline științifice importante, cum ar fi, de pildă,programarea liniară. Este de notat, însă, că toate aceste aplicații, de oricare natură ar fi ele, se fundamentează pe statistică, astfel că putem întrevedea că progresul realizat în tehnica calculului va da un mare impuls științei statistice.

În legătură cu aceasta este important de subliniat că în U.R.S.S. funcționează centre de calcul, înzestrate cu mașini electronice, care lucrează exclusiv pentru economie.

Mașinile electronice de calcul încă de pe acum au deschis mari posibilități de aplicare a statisticii, care din ce în ce mai mult devine o metodă de cercetare indispensabilă în economie, știință și tehnică.

Pentru a răspunde complet la întrebarea pusă, vom adăuga că, cel puțin în actuala fază de construire a mașinilor electronice de calcul, ele nu vor contribui la diminuarea importanței metodei selective, ci dimpotrivă la o creștere.

Mașinile electronice aduc mari avantaje în prelucrarea datelor, însă, cel puțin pînă în prezent, nu fac același lucru pentru culegerea lor. Acest fapt constituie un impediment pentru prelucrarea totală. Mai există apoi cazuri cînd prelucrarea totală, după cum am mai arătat, aduce după sine pierderea totală sau parțială a materialului selecționat. De exemplu, determinarea recoltei probabile în agricultură nu va putea fi făcută, din motive lesne de înțeles, folosind întreaga suprafață cultivată. Toate

*) A. Nesmeianov, *Nauka i stroitelstvo kommunizma*, Komunist, *1* (1959).

aceste argumente arată clar că selecția va rămîne una din metodele fundamentale ale practicii statisticii.

Pe de altă parte, aplicarea pe scară din ce în ce mai largă a statisticii în cele mai diverse ramuri de activitate, va duce la o dezvoltare progresivă a matematicilor aplicate în cercetarea statistică și implicit, la dezvoltarea teoretică a metodei selecției, care constituie baza aplicării matematicii în statistică.

2. Reprezentativitatea selecției

Principiul selecției formulat mai sus învederează, în esență, că valoarea cognitivă a metodei selecției depinde în mod hotărîtor de modul în care a fost aleasă colectivitatea de selecție. Pentru aceasta, este nevoie ca partea selecționată din colectivitatea generală să fie cît mai *reprezentativă* și în acest scop se impune — după cum am mai arătat — ca elementele extrase să aibă o egală probabilitate de a face parte din selecția obținută. În plus, volumul selecției obținute trebuie să fie destul de mare, iar elementele din care este formată colectivitatea generală să fie cît mai omogene.

Numai în aceste condiții, valorile caracteristicilor urmărite vor diferi foarte puțin de aceleași valori ale colectivității generale, așa încît rezultatele obținute cu ajutorul selecției se vor putea extinde asupra întregii colectivități.

Reprezentativitatea selecției se asigură cînd structura colectivității de selecție este identică cu structura colectivității generale sau diferă foarte puțin de aceasta. În cazul unei identități depline a structurii, se spune că avem o reprezentativitate perfectă. O asemenea reprezentativitate, însă, nu se poate realiza practic, decît cu totul întîmplător. De obicei, se consideră satisfăcătoare acea selecție care asigură abateri ale indicatorilor de structură ai colectivității de selecție cu cel mult 5%, față de indicatorii corespunzători din colectivitatea generală.

În cazul cercetărilor selective continue, cum sînt cercetările bugetelor de familie ale muncitorilor și colectiviștilor, se pune problema verificării periodice a reprezentativității selecției. În U.R.S.S., reprezentativitatea selecției prin care se cercetează bugetele colhoznicilor se verifică periodic; aceeași verificare se face și la noi. În acest scop se calculează indicatorii sintetici, cum sînt media și dispersia, pentru caracteristicile studiate din colectivitatea de selecție și se compară cu indicatorii respectivi din colectivitatea generală. Dacă diferențele sînt mici, adică se află în limitele unui interval probabilistic dinainte stabilit, aceasta înseamnă că selecția continuă să fie reprezentativă; în caz contrar, selecția trebuie refăcută.

Este evident că prin metoda selecției se vor produce unele erori, dintre care cele mai importante sînt:

a) *Erori de înregistrare.* Ca și la înregistrările totale, aceste erori se produc adeseori din cauza unor factori cu totul întîmplători care intervin în procesul culegerii datelor și îndeosebi din cauza neatenției celor însărcinați cu efectuarea acestei operații; este vorba de erori nesistematice. Alteori,

asemenea erori pot fi, însă, *sistematice* (premeditate sau nepremeditate), ca, de pildă, în cazul cînd cei însărcinați cu completarea datelor, în tendința lor de a înfrumuseța situația reală sau de a ascunde unele lipsuri, înregistrează intenționat anumite date inexacte.

În ultima analiză, erorile de înregistrare nesistematice se compensează reciproc, așa încît exercită, de obicei, o influență minimă, nesemnificativă, asupra exactității datelor.

b) *Erori de reprezentativitate.* Asemenea erori au loc cînd nu se asigură tuturor elementelor din colectivitatea generală aceeași probabilitate să ajungă în selecție. Ele sînt de două feluri : erori *sistematice* și *erori întîmplătoare.*

Prin erori de reprezentativitate se înțeleg *diferențele dintre nivelurile caracteristicilor studiate din colectivitatea de selecție și cele din colectivitatea generală.* Aceste diferențe se produc în toate cazurile, deoarece selecția nu *reprezintă* exact întreaga colectivitate.

Eliminarea completă a erorilor de reprezentativitate nu este posibilă deoarece aceasta ar însemna să înlocuim observarea parțială prin observarea totală. În schimb, este posibil calculul cu exactitate al mărimii *probabile* a acestor erori, ceea ce face ca metoda selecției să fie un instrument foarte puternic de cercetare, ocupînd un loc deosebit de important în statistica matematică și, implicit, în diversele aplicații ale acesteia, printre care și în cercetarea fenomenelor social-economice.

Ca să sesizăm mai bine esența erorilor de reprezentare vom considera următorul exemplu : Să presupunem că populația unui oraș este formată din N persoane și că ne propunem să determinăm mortalitatea populației acestui oraș, folosind o selecție de volum $n < N$. Dacă selecția noastră este formată numai din persoane de 70 ani, este clar că am făcut o gravă eroare de înregistrare, deoarece mortalitatea acestui grup de persoane este evident mai mare decît mortalitatea întregii populații. Eroarea de reprezentativitate va fi minimă dacă structura pe vîrste a locuitorilor din selecție va fi aceeași cu structura pe vîrste a locuitorilor din întregul oraș. Acest lucru se presupune că se obține cînd, n fiind destul de mare, se face o selecție a locuitorilor la întîmplare.

După cum se vede, vom fi feriți de erori de reprezentativitate cînd structura selecției este apropiată de structura colectivității generale, adică atunci cînd selecția este reprezentativă.

Erorile de reprezentare *sistematice* sînt efectul nerespectării principiului selecției pe care l-am formulat mai sus, încălcîndu-se regula ca fiecărei unități din colectivitatea generală să-i fie asigurată aceeași probabilitate de a intra în selecție.

Aceasta are loc, de pildă, în cazul cînd — la cercetarea recoltei medii la hectar — probele cu rama metrică se iau, de preferință, sau dintr-o greșită înțelegere a executantului, din părțile lanului cu recoltă bună. Este evident că în acest caz selecția nu va fi reprezentativă, întrucît greutatea specifică a probelor cu recolta bună va fi mai mare în selecție decît în întregul lan. Aceste erori pot avea loc și din multe alte cauze, printre care și din cauza delimitării defectuoase a unităților care trebuie incluse în selecție.

Scanned with OKEN Scanner

Erorile de reprezentativitate *întîmplătoare* sînt de esența metodei însăși. Ele au loc chiar și în ipoteza că selecția s-a executat cu respectarea riguroasă a principiului selecției, dat fiind că, și în acest caz întîmplător, structura selecției poate diferi de structura colectivității generale. Pentru ca selecția să poată oferi rezultate suficient de reprezentative este absolut necesar ca erorile întîmplătoare să fie cît se poate de mici. Ordinul de mărime a acestor erori depinde de volumul selecției, de variabilitatea materialului observat, de metoda de selecție folosită, ca și de metoda de prelucrare și de calcul adoptată. Statistica matematică a elaborat metode prin care mărimea erorii întîmplătoare poate fi stabilită anticipat. Aceasta constituie o mare cucerire a științei, a metodelor elaborate îndeosebi de statistica matematică. Cu ajutorul metodelor statisticii matematice, se poate stabili chiar și precizia relativă a diferitelor selecții, operate asupra unui material dat, ca și planurile unor cercetări selective viitoare.

Tocmai descoperirea acestor metode a făcut posibilă dezvoltarea calitativă a metodei selecției, inițial îndoielnică și nesigură. Metoda selecției a devenit în prezent o metodă riguros definită și bine determinată, căreia i se poate acorda toată încrederea. Mai mult, posibilitatea pe care o avem de a fixa limitele erorii întîmplătoare de reprezentativitate ne permite să scoatem în evidență celelalte tipuri de erori amintite (provenind dintr-o selecție prost executată sau de la metode de observare greșit alese), precum și erori care afectează alte surse de date statistice cu care pot fi comparate rezultatele selecției.

Pentru eliminarea sau reducerea erorilor de reprezentativitate întîmplătoare se pot folosi mai multe metode, printre care, în primul rînd — o metodă cu totul simplă — aceea care se bazează pe sporirea volumului selecției. Toate celelalte condiții rămînînd neschimbate, în prima aproximare, eroarea întîmplătoare este în general invers proporțională cu rădăcina patrată din numărul unităților incluse în selecție.

Precizia selecției depinde totodată nu numai de numărul unităților cuprinse în selecție ci și de variația caracteristicii de la o unitate la alta sau, mai exact, de acea parte a variației care contribuie la apariția erorii întîmplătoare. Tocmai aici se prezintă complicațiile tehnice ale selecției și totodată dificultățile legate de alegerea planului ei și de analiza rezultatelor obținute. Datorită procedeelor potrivite de selecție care impun diferite restricții și se fundamentează pe o selecție făcută exclusiv la întîmplare, variația între unități care determină erorile întîmplătoare poate fi sensibil redusă și, în consecință, volumul selecției necesar pentru a avea o precizie dată poate fi și el redus.

Tipul cel mai simplu al unor restricții de acest fel este acela cunoscut sub numele *grupare* sau *stratificare* cum i se mai spune. Colectivitatea este grupată (stratificată) în așa fel ca în fiecare grupă (strat) tipică să fie întrunite unități cît mai omogene, respectîndu-se, deci, regulile cunoscute ale grupării după una sau mai multe caracteristici.

Din fiecare grupă tipică în care a fost despărțită colectivitatea generală se extrage apoi la întînplare un anumit număr de unități, reprezentînd un procent bine determinat. Dacă din fiecare grupă se extrage un

Scanned with OKEN Scanner

număr de unități proporțional cu volumul grupei, este clar că fiecare grupă va fi corect reprezentată și, în consecință, deosebirile dintre diferite grupe în ce privește erorile de reprezentativitate vor fi eliminate.

În afară de operația de grupare mai sînt și alte procedee care vor fi discutate amănunțit mai jos, datorită cărora se poate asigura selecției o precizie mai mare. Este vorba, de pildă, despre procedeul utilizării informațiilor suplimentare, despre folosirea unei fracțiuni de selecție variabilă (denumită adeseori *selecție optimă*) ca și despre selecția în faze.

Utilizarea informațiilor suplimentare, în general, se referă la date din alte surse decît cele utilizate pentru selecție sau la date obținute pe baza unei alte selecții de volum mai mare decît al selecției efectuate.

Folosirea unei fracțiuni variabile de selecție înseamă reprezentarea neproporțională a diferitelor grupe în selecție. Acest procedeu permite ca selecția să se efectueze mai intens în acele părți ale colectivității generale care sînt mai importante și în care variația caracteristicilor studiate este mai mare. Dacă se aplică procedeul cotelor părți neproporționale este absolut necesar să se pondereze corect participarea fiecărei grupe în totalul colectivității studiate,

Selecțiile optime depind de variația relativă a diferitelor grupe în care a fost despărțită colectivitatea în vederea alegerii eșantionului. Astfel, dacă ne propunem de a estima numărul muncitorilor într-o ramură industrială oarecare, va fi potrivit ca o fracțiune de selecții cît mai mare să fie atribuită întreprinderilor mari.

În selecții *cu mai multe faze*, colectivitatea se împarte mai întîi într-un anumit număr de grupe k din care, în prima fază, se extrage un număr $h < k$ grupe și apoi din cele h grupe se extrag la întîmplare altele $l < h$ grupe etc. Un asemenea procedeu asigură, evident, un grad mai mare de precizie a selecției.

3. Cercetarea selectivă în lucrările de recensăminte și anchete

Metoda selecției este folosită adeseori în practica recensămintelor. Să presupunem, astfel, că se urmărește determinarea unei caracteristici care ar putea fi obținută pe baza unui recensămînt general. Să presupunem că trebuie să se determine numărul vitelor dintr-o anumită regiune, folosite în agricultură. Pentru realizarea acestui scop se folosește adeseori un recensămînt parțial, considerîndu-se la întîmplare numai 5% sau 10% din gospodăriile agricole din regiune. Se ajunge la rezultat pe cale mai rapidă, mai puțin costisitoare și de multe ori chiar mai exactă pentru că reducerea materialului statistic de cules la proporții mai restrînse oferă posibilitatea unei cercetări statistice mai precise.

Trebuie să arătăm, însă, că selecțiile pot fi utilizate în lucrările de recensămînt numai cu condiția ca erorile de reprezentativitate să fie atît de mici încît, ținînd seama de scopul urmărit, valabilitatea rezultatelor obținute prin selecție să nu fie afectată.

Aceasta va depinde, în particular, de gradul de detaliere a rezultatelor scontate. Dacă ne interesează numai rezultatele cu privire la colectivi-

Scanned with OKEN Scanner

tatea generală, o precizie dată a selecției poate fi obținută cu ajutorul unei selecții cu mult mai mici decît dacă ne-ar interesa rezultatele de detaliu, pentru diferite părți ale colectivității generale (în limitele unei regiuni, oraș etc.). În unele cazuri, volumul selecției trebuie să fie atît de mare încît se pune practic problema de a ști dacă nu este mai preferabilă efectuarea unui recensămînt general. Evident că în caz extrem — dacă avem nevoie de date asupra fiecărei unități, efectuarea unui recensămînt general nu poate fi evitată.

Un alt factor de care trebuie să ținem seama cînd ni se pune problema de a decide dacă trebuie efectuată o selecție sau un recensămînt general este acela al dificultăților pe care le implică precum și al costului pe care-l reclamă o operație sau alta. Trebuie să ținem seama cu acest prilej că volumul de muncă și cheltuielile pe unitate de observare pe care le necesită culegerea datelor sînt în general mai mari în cazul selecției decît în cazul recensămîntului complet. Mai mult, selecția are problemele ei proprii de organizare, din care unele nici nu se pun la organizarea unui recensămînt complet; se poate întîmpla chiar ca — în cazul unui program de observare simplu conceput — recensămîntul să se poată efectua exclusiv prin munca organelor locale ale puterii de stat, pe cînd o selecție ar reclama punerea în funcțiune a unui aparat complex, special organizat. Totuși, în general, dacă selecția necesară nu reprezintă decît o fracțiune mică a colectivității totale, volumul de muncă și de cheltuieli are să fie cu mult mai mic în cazul cercetării selective decît în cazul efectuării unui recensămînt general.

De aceea, în majoritatea cazurilor, aplicarea metodei selecției aduce o mare economie de timp.

Mai sînt și o serie de alte avantaje care nu se pot remarca imediat. În primul rînd, în cazul cercetării selective, răspunsurile la întrebările din programul observării se obțin mult mai ușor, mai complet și mai exact, datorită volumului mic al unităților observate.

În al doilea rînd, culegerea datelor se poate face după un program mai amănunțit, obținîndu-se date multilaterale asupra obiectului cercetat, iar în faza prelucrării și analizei acestora se pot executa lucrări de calitate mai bună, datorită faptului că este posibilă alegerea cadrelor mai bine pregătite, specializate (volumul lucrărilor fiind mic) și că în condițiile aceluiași timp consacrat diferitelor calcule poate fi operat un număr mai mare de grupări din cele mai variate.

În al treilea rînd, selecția accelerează considerabil încheierea diferitelor etape de cercetare, începînd cu etapa inițială de culegere a datelor și terminînd cu analiza statistică a rezultatelor obținute.

Este evident că metoda selecției poate fi folosită și pentru prelucrarea sau analiza materialelor obținute printr-o înregistrare totală, aceasta cu scopul de a satisface cerințele unei conduceri operative, imediate (așa s-a procedat la noi în țară, cînd s-a pus problema prelucrării rapide a datelor recensămîntului agricol din 1948).

Metoda selecției se utilizează uneori paralel cu înregistrarea totală sau chiar premerge acesteia din urmă, așa cum s-a procedat în U.R.S.S., înainte de efectuarea recensămîntului populației de la 1 ianuarie 1960.

Scanned with OKEN Scanner

În statistica socialistă, metoda selecției este folosită uneori și ca mijloc de cercetare a exactității datelor obținute prin înregistrarea totală (de pildă, recensămîntul anual al animalelor este urmat de o anchetă selectivă cuprinzînd, de obicei, un volum de cel puțin 10% din numărul gospodăriilor cercetate prin observarea totală).

În etapa actuală de dezvoltare a statisticii socialiste — implicit a statisticii din țara noastră — metoda selecției își găsește aplicații, pe scară tot mai largă, în cercetarea fenomenelor și proceselor care au loc în cadrul larg al reproducției socialiste lărgite.

4. Formularea matematică a principiului selecției

Vom preciza mai întîi sensul unor categorii statistice și notația pe care o vom folosi în această parte a lucrării, fără să ne oprim asupra unor noțiuni pe care cititorul le-a mai întîlnit în cursurile de statistică teoretică și fără a mai schimba o serie de simboluri folosite în partea în care a fost expus calculul probabilităților.

Prin *colectivitate generală* — categorie folosită și pînă acum — vom înțelege orice colectivitate statistică — masă de unități elementare ale fenomenului studiat — asupra căreia se extind rezultatele selecției. În practică avem de a face cu colectivități generale finite.

Unitățile pe care le extragem din colectivitatea generală vor fi denumite *unități de selecție* care pot fi ori unități elementare ori grupe de unități elementare; totalitatea unităților selectate va fi denumită *colectivitatea de selecție*. Concluziile pe care le tragem cu privire la nivelurile pe care le constatăm la diferite caracteristici în colectivitatea de selecție se extind asupra colectivității generale.

Vom introduce și următoarea notație *):

X sau X_i, simbolul principal pentru o caracteristică cantitativă, variabilă, pe care o posedă unitățile din colectivitatea generală. (Pentru cazul cu două sau mai multe caracteristici se vor introduce și alte litere majuscule Y, Z, sau A, B etc);

x și x_i -idem — pentru caracteristicile unităților din colectivitatea de selecție (sau y, z etc.).

Dacă colectivitatea generală sau colectivitatea de selecție se împarte în grupe (straturi) tipice, atunci vom folosi un simbol suplimentar $k(X_k, X_{ki}; x_k, x_{ki})$ sau h.

Diferite grupe tipice ale colectivității se vor nota cu $k_i (i = 1,2,..,k)$, iar simbolurile referitoare la aceste grupe se vor nota cu semne suplimentare (de exemplu, n_{k_1}, n_{k_2}, n_{k_3}, . . .).

Notațiile specifice pentru o serie de alți indicatori vor fi introduse pe parcurs odată cu expunerea diferitelor probleme.

★

*) Precizăm că pentru caracteristicile unităților din colectivitatea generală vom folosi de obicei litere latine majuscule.

Să considerăm, de pildă, producția în masă a unor piese, executată cu ajutorul unei mașini. Piesele — cele acceptabile ca și cele neacceptabile (sub raportul îndeplinirii anumitor condiții de calitate) — formează o colectivitate statistică pe care o supunem studiului din anumite puncte de vedere. Să presupunem în acest scop că — în condițiile normale de funcționare a mașinii — ne este cunoscută legea de probabilitate a distribuției teoretice a acestor piese.

Cunoscînd legea de probabilitate a distribuției pieselor, sîntem în măsură de a determina corect valoarea medie a numărului de piese acceptabile produse într-un interval de timp dat, împrăștierea lor în jurul valorii medii etc. De asemenea, sîntem în măsură de a sezisa momentul cînd mașina încetează de a mai funcționa în condiții normale. Mai precizăm că uneori, din punct de vedere practic, nu ne interesează decît anumiți indicatori statistici de sinteză, cum ar fi, de exemplu primele patru momente centrate.

Ne aflăm, după cum se vede, într-o ipoteză cu totul favorabilă pentru a determina indicatorii statistici ce ne interesează.

Un interes deosebit îl prezintă, însă, o altă ipoteză, și anume, cînd legea de probabilitate a fenomenului studiat nu ne este cunoscută dinainte și nici nu o putem determina a priori.

Problema care se pune în acest caz — și trebuie să subliniem că ea constituie una din problemele fundamentale ale statisticii matematice — este aceea de a găsi legea de probabilitate a distribuției colectivității generale sau de a determina cel puțin indicatorii principali care ne interesează și aceasta numai *pe baza datelor reale pe care ni le oferă observarea parțială a fenomenului.*

Vom arăta că, în anumite condiții, rezolvarea acestei probleme este posibilă și că în acest scop este folosită tocmai *metoda selecției.*

Formulată în termeni matematici, esența acestei metode este următoarea: fiind dată o colectivitate formată din N elemente, în urma a n experiențe independente, bine determinate, obținem rezultatele $x_1, x_2, \ldots, x_n$, privind o caracteristică oarecare X, luată în studiu. Se pune problema de a fi determinați indicatorii sintetici din colectivitatea generală, cum sînt momentele, mediana etc., cu ajutorul variabilelor x_i, $(i = 1,2,\ldots, n)$.

Variabilele x_i sînt *variabile de selecție* iar o funcție de tipul $f(x_1, x_2, \ldots, x_n)$ este o *funcție de selecție.*

Vom avea astfel de considerat diferite funcții de selecție, ca de exemplu, media de selecție

$$\bar{x} = \frac{x_1 + x_2 + \ldots + x_n}{n}$$

momentul de selecție de ordinul r

$$\bar{M}_r = \frac{x_1^r + x_2^r + \ldots + x_n^r}{n}$$

și așa mai departe.

O deosebită importanță o prezintă cazul cînd funcția de selecție tinde în probabilitate către o limită bine determinată.

Cea mai simplă funcție de selecție este funcția empirică de selecție sau funcția de repartiție de selecție.

Fie $F(x)$ funcția teoretică de repartiție. În problemele întîlnite de obicei în statistică, am văzut că această funcție este sau continuă sau o funcție cu trepte, corespunzător celor două distribuții clasice din statistică: distribuția continuă și distribuția discontinuă. Se pune în mod natural problema convergenței funcției empirice de selecție către funcția teoretică de repartiție. Cel dintîi care a adus o contribuție importantă în rezolvarea acestei probleme a fost Cantelli *), care a tratat cazul urnei lui Bernoulli.

Fie p probabilitatea constantă de a extrage o bilă albă și $f_n = \frac{\alpha}{n}$, frecvența corespunzătoare în n extracții succesive. α reprezintă numărul de bile albe obținute în n extracții. În baza teoremei lui Bernoulli, putem afirma că f_n tinde în probabilitate către p, cînd $n \to \infty$.

Să considerăm acum șirul infinit

$$f_n - p,\ f_{n+1} - p,\ f_{n+2} - p,\ \ldots \qquad \text{(XIX, 4.1)}$$

Ne interesează, din punct de vedere practic, probabilitatea ca acest șir să tindă către zero. Cantelli a demonstrat că probabilitatea ca acest șir să tindă către zero este 1. Este de observat că numai convergența în probabilitate a frecvenței f_n către p nu este suficientă pentru a deduce teorema lui Cantelli. Probabilitatea evenimentului $|f_n - p| > \varepsilon$, pentru $n > N$, poate fi făcută atît de mică cît dorim. Nu putem fi siguri însă că dacă vom considera evenimentele în număr infinit

$$|f_{N+1} - p| > \varepsilon, \quad |f_{N+2} - p| > \varepsilon, \ldots, |f_{N+s} - p| > \varepsilon, \quad (s = 1,2, \ldots)$$

nu se întîmplă ca unul din ele să se producă, cu toate că fiecare are o probabilitate foarte mică

Demonstrația teoremei lui Cantelli se găsește expusă la capitolul VII unde ne-am ocupat de legea numerelor mari.

Convergența tare a unui șir de variabile aleatoare către zero

$$A_1, A_2, \ldots, A_n, \ldots$$

se poate scrie

$$P(A_n \to 0) = 1.$$

În cazul urnei lui Bernoulli, putem scrie

$$P\left(\frac{\alpha}{n} \to p\right) = 1.$$

Această proprietate de limită constituie teorema lui Cantelli. Să admitem că am efectuat o selecție de mărime n, obținînd rezultatele

$$x_1, x_2, \ldots, x_n \qquad \text{(XIX, 4.2)}$$

Funcția de repartiție $F_n(x)$ a șirului (XIX, 4.2) așezînd în fiecare x_i, $(i = 1, 2, \ldots, n)$, masa $\frac{1}{n}$, poartă numele de funcție empirică de selecție.

*) F. P. Cantelli, *Sulla probabilità come limite della frecuenza.* Réndiconte della Accademia dei Lincei, serie V, XXVI (1917).

Următoarea teoremă a lui Glivenko (1933), numită de unii autori teorema fundamentală a statisticii matematice, dă justificarea teoretică a folosirii metodei selecției.

TEOREMA LUI GLIVENKO*). *Notînd prin $F(x)$ funcția de repartiție a unei distribuții teoretice și prin $F_n(x)$ funcția empirică de repartiție, obținută într-o selecție de mărime n, avem:*

$$P[\sup_{-\infty<x<\infty} |F_n(x) - F(x)| \to 0] = 1.$$

Demonstrația care urmează aparține lui B. V. Gnedenko **). Ea se caracterizează prin simplitate și eleganță.

Pentru demonstrarea teoremei sînt necesare următoarele două leme:

LEMA 1. *Dacă evenimentul E constînd din realizarea simultană a unui număr infinit de evenimente*

$$E_1, E_2, \ldots, E_n, \ldots$$
$$E = E_1 E_2, \ldots$$

astfel încît realizarea fiecărui eveniment E_{n+1} implică realizarea precedentului său, avem

$$P(E) = \lim_{n\to\infty} P(E_n).$$

LEMA 2. *Dacă probabilitatea fiecărui eveniment E_i, $(i = 1, 2, \ldots)$, este egală cu 1, avem de asemenea $P(E) = 1$.*

Demonstrația acestor leme este simplă.

Putem scrie, notînd prin $\bar{F}$ evenimentul opus lui F:

$$\begin{aligned} E_1 &= E_1\bar{E}_2 + E_2\bar{E}_3 + \ldots + E_{n-1}\bar{E}_n + E_n, \\ E_1 &= E_1\bar{E}_2 + E_2\bar{E}_3 + \ldots + E_{n-1}\bar{E}_n + E_n\bar{E}_{n+1} + \ldots + E. \end{aligned} \qquad \text{(XIX, 4.3)}$$

Aceste relații sînt evidente dacă ținem seama de faptul că evenimentul E_{n+1} implică E_n.

Evenimentele care sînt date de termenii din membrul al doilea ai relațiilor (XIX, 4.3) sînt incompatibile între ele. Rezultă prin urmare:

$$\begin{aligned} P(E_1) &= P(E_1\bar{E}_2) + P(E_2\bar{E}_3) + \ldots + P(E_{n-1}\bar{E}_n) + P(E_n) \\ P(E_1) &= P(E_1\bar{E}_2) + P(E_2\bar{E}_3) + \ldots + P(E_{n-1}\bar{E}_n) + \\ &\quad + P(E_n\bar{E}_{n+1}) + \ldots + P(E) \end{aligned} \qquad \text{(XIX, 4.4)}$$

*) V. I. Glivenko (1897–1940), matematician sovietic. În primele sale lucrări s-a ocupat de bazele matematicii și de logica matematică, abordînd primul bazele matematicii de pe pozițiile materialismului dialectic. În statistica matematică, Glivenko a demonstrat teorema tendinței de apropiere a legii empirice de distribuție către legea teoretică pe care o dăm mai sus. În ultimii ani Glivenko s-a ocupat de teoria structurilor.

**) B. V. Gnedenko, *Curs teoria veroiatnostei*. Moskva, 1954, p. 327–331.

Scanned with OKEN Scanner

Scăzînd prima relația din a doua obținem mai departe:

$$P(E) = P(E_n) - \sum_{k=n}^{\infty} P(E_k \overline{E}_{k+1}).$$

Însă în baza ultimei relații din (XIX, 4.4) avem

$$\lim_{n \to \infty} \sum_{k=n}^{\infty} P(E_k \overline{E}_{k+1}) = 0.$$

Rezultă astfel

$$P(E) = \lim_{n \to \infty} P(E_n).$$

Am demonstrat astfel lema 1.

Pentru un șir finit de variabile E_i, lema 2 rezultă imediat. Astfel, pentru două evenimente E_1 și E_2 să presupunem că

$$P(E_1) = 1, \qquad P(E_2) = 1.$$

Știm că

$$P(E_1) = 1, \qquad P(E_2) = 1, \qquad P(E_1 + E_2) = 1.$$

Însă

$$P(E_1 + E_2) = P(E_1) + P(E_2) - P(E_1 E_2).$$

Rezultă imediat

$$P(E_1 E_2) = 1.$$

Procedînd în mod analog se demonstrează:

$$P(E_1 E_2 \ldots E_n) = 1.$$

Putem scrie

$$E_1 E_2 E_3 \ldots = E_1 (E_1 E_2)(E_1 E_2 E_3) \ldots$$

Fiecare factor din membrul al doilea implică precedentul, astfel încît în baza lemei 1 putem scrie

$$P(E_1 E_2 \ldots) = \lim_{n \to \infty} P(E_1 E_2 \ldots E_n).$$

Rezultă

$$P(E_1 E_2 \ldots) = 1,$$

ceea ce demonstrează lema 2.

Putem acum trece la demonstrația teoremei lui Glivenko. Fie $r > 0$ un număr întreg și $x_{r,k}$ cel mai mic x, care satisface inegalitățile

$$F(x - 0) = F(x) \leqslant \frac{k}{r} < F(x + 0) \ldots, \quad (k = 1, 2, \ldots, r). \quad \text{(XIX, 4.5)}$$

Dacă notăm prin ξ variabila aleatoare care are ca funcție de repartiție pe $F(x)$, din inegalitățile precedente rezultă

$$P(\xi < x_{r,k}) = F(x_{r,k}).$$

În baza teoremei lui Cantelli putem deci scrie

$$P[\lim_{n \to \infty} F_n(x_{r,k}) \to F(x_{r,k})] = 1, \quad (k = 1, 2, \ldots, r). \quad \text{(XIX, 4.6)}$$

Să notăm prin E_k^r evenimentul de mai sus, adică pentru care

$$\underset{n\to\infty}{F_n}(x_{r,k}) \to F(x_{r,k}), \qquad (k = 1, 2, \ldots, r).$$

În baza relației (XIX, 4.6) avem:

$$P(E_k^r) = 1, \qquad (k = 1, 2, \ldots, r). \qquad \text{(XIX, 4.7)}$$

Să considerăm acum și evenimentul

$$E^r = E_1^r E_2^r \ldots E_r^r$$

care poate fi scris sub forma

$$\max_{1\leqslant k\leqslant r_1} \underset{n\to\infty}{|F_n(x_{r,k}) - F(x_{r,k})|} \to 0.$$

În baza lemei 2 și relației (XIX, 4.7) rezultă că

$$P(E^r) = 1. \qquad \text{(XIX, 4.8)}$$

Să notăm acum prin E evenimentul $E^1E^2E^3\ldots$ și prin S evenimentul

$$\sup |F_n(x) - F(x)| \to 0$$

oricare ar fi x pe axa reală.

În baza lemei 2 și relației (XIX, 4.8) avem

$$P(E) = 1. \qquad \text{(XIX, 4.9)}$$

Dacă x este cuprins între $x_{r,k}$ și $x_{r,k+1}$, există inegalitățile

$$F_n(x_{r,k} + 0) \leqslant F_n(x) \leqslant F_n(x_{r,k+1}),$$
$$F(x_{r,k} + 0) \leqslant F(x) \leqslant F(x_{r,k+1}),$$
$$0 \leqslant F(x_{r,k+1}) - F(x_{r,k} + 0) \leqslant \frac{1}{r}.$$

Prin urmare

$$F_n(x_{r,k} + 0) - F(x_{r,k+1}) \leqslant F_n(x) - F(x) \leqslant F_n(x_{r,k+1}) - F(x_{r,k} + 0)$$

sau

$$|F_n(x) - F(x)| \leqslant \max_{1\leqslant k\leqslant r} |F_n(x_{r,k}) - F(x_{r,k})| + \frac{1}{r}$$

și

$$\sup_{-\infty < x < \infty} |F_n(x) - F(x)| \leqslant \max_{1\leqslant k\leqslant r} |F_n(x_{r,k}) - F(x_{r,k})| + \frac{1}{r}.$$

Cum r este arbitrar inegalitatea de mai sus ne arată că E este inclus în S și cum $P(E) = 1$, rezultă și $P(S) = 1$, adică

$$P[\sup_{-\infty < x < \infty} |F_n(x) - F(x)| \to 0] = 1,$$

adică teorema lui Glivenko *).

*) O demonstrație care nu diferă esențial de aceasta se găsește în tratatul: O. Onicescu și Gh. Mihoc, *Lecții de statistică matematică*, Ed. tehnică, București, 1958, p. 662–664.

Scanned with OKEN Scanner

Capitolul XX

Structura diferitelor tipuri de selecţie. Sondaje cu o fază

Pentru simplificare, în cele ce urmează subcolectivitatea de selecţie o vom denumi sondaj.

1. Consideraţii generale

Vom expune, în primul rînd, problema sondajelor care se întîlnesc cel mai frecvent în practica muncii de statistică şi anume a sondajelor pe care le extragem din colectivităţi finite. Atragem atenţia, însă, că problema se reduce uneori la acea a sondajelor dintr-o colectivitate infinită. Vom considera, de aceea, după caz, fie schema probabilistică clasică a urnei lui Bernoulli cu bila revenită, fie schema teoretică hipergeometrică.

În primul caz vom avea o *selecţie repetată*, cînd fiecare unitate a colectivităţii generale, ipotetic extrasă, rămîne de fapt în cadrul acestei colectivităţi. În modul acesta, variabilele selecţiei pot fi independente între ele, iar fiecare din unităţile colectivităţii poate fi extrasă de mai multe ori, în acest caz, variabilele selecţiei dau naştere unei distribuţii teoretice după modelul urnei lui Bernoulli. De altfel, procedînd astfel după schema urnei lui Bernoulli, succesiunea probelor de selecţie este infinită.

În al doilea caz, vom avea o *selecţie nerepetată* concretizată printr-o urnă din care facem extracţii succesive, fără să punem bila înapoi în urnă; variabilele selecţiei sînt dependente între ele, iar o unitate nu poate să apară în şirul succesiv al selecţiilor decît o singură dată. Avem, deci, o selecţie dintr-o colectivitate finită.

Unele condiţii de ordin practic fac posibilă înţelegerea mai bună a deosebirilor dintre diferitele feluri de sondaje.

Să presupunem că ne interesează procentul de analfabeţi dintr-o populaţie, pe baza datelor unui recensămînt general.

Aplicînd metoda selecției, prelucrăm un material de 5%, luat la întîmplare, din volumul total. Dacă în volumul selecției vom alege numai locuitorii din mediul urban, rezultatul obținut ar fi profund eronat, în cazul cînd în realitate există o diferență mai mare între procentele de analfabeți din mediul urban și cel rural. La fel am ajunge la un rezultat eronat dacă tot sondajul ar fi efectuat numai din populația rurală.

Procentul de 5% din populația totală care ne dă volumul sondajului va trebui luat și din populația urbană și din populația rurală. Vom aplica deci metodologia sondajului pe grupe (stratificat) de care ne vom ocupa mai jos.

Mai trebuie precizată aici noțiunea de *eficiență* a unui sondaj.

Vom spune că un sondaj A de mărimea n, în baza căruia estimăm $\overline{X}$ prin funcția de selecție $\overline{x}_A$ este mai *eficient* decît sondajul B, tot de mărimea n în baza căruia estimăm indicatorul $\overline{X}$ prin funcția de selecție $\overline{x}_B$, dacă avem relațiile

$$M(\overline{x}_A) = \overline{X}, \qquad M(\overline{x}_B) = \overline{X}, \qquad D^2(\overline{x}_A) < D^2(\overline{x}_B).$$

Se vede imediat de ce sondajul A este preferabil sondajului B.

Cu o probabilitate mai mică decît $1 - \frac{1}{k^2}$ putem afirma, în cazul sondajului A, că media teoretică se găsește cuprinsă în intervalul $[\overline{X} - kD(\overline{x}_A),\ \overline{X} + kD(\overline{x}_A)]$ iar în cazul sondajului B în intervalul $[\overline{X} - kD(\overline{x}_B),\ \overline{X} + kD(\overline{x}_B)]$.

Deci, cu același prag de probabilitate, găsim pentru media teoretică un interval mai mic de acoperire în cazul sondajului A decît în cazul sondajului B, ceea ce în mod evident constituie un avantaj.

2. Sondajul simplu nerepetat

1°. *Cazul cînd este studiată o singură caracteristică.* Sondajul simplu presupune că subcolectivitatea de selecție — sondajul — se alege la întîmplare, din întreaga colectivitate generală. Precizia acestei selecții depinde atît de variația caracteristicii studiate în cadrul colectivității generale cît și de volumul n al selecției. Sensul logic al acestor condiții este evident. Cu cît caracteristica studiată are variații mai mari de la o unitate la alta a colectivității generale, cu atît mai greu este să reproduci dimensiunile ei exacte în selecție și, de aceea, cu atît mai mare se impune volumul selecției pentru a atinge acest scop. Așa, de pildă, cînd determinăm recolta medie la hectar, volumul selecției depinde de gradul de variație a recoltei. Dacă la toate lanurile recolta este aproximativ aceeași este suficient să luăm *un singur* lan și, pe baza datelor privind recolta medie la hectar, să apreciem recolta medie la hectar la toate lanurile. În situația inversă, cînd recolta are variații mari de la un lan la altul, volumul selecției trebuie mărit considerabil.

Scanned with OKEN Scanner

Această ultimă ipoteză presupune necesitatea de a fi stabilită relația matematică între indicatorii variației și indicatorul volumului selecției.

Un sondaj simplu are de obicei drept scop imediat *estimarea mediei* unor caracteristici cantitative dintr-o colectivitate statistică generală sau estimarea *frecvenței relative* a unităților din colectivitatea generală, purtătoare ale unei caracteristici calitative.

Vom examina modul în care se poate atinge acest scop și fundamentarea lui matematică.

Să considerăm colectivitatea A, ale cărei unități prezintă fiecare o singură caracteristică. Unitățile sondajului au, respectiv variantele unor anumite caracteristici $a_1, a_2, \ldots, a_n$. De exemplu, dacă avem o colectivitate formată din N persoane salariate și ne interesează salariul lor, numerele $a_1, a_2, \ldots, a_N$ reprezintă salariul fiecărei persoane.

Știm că există diferite valori tipice care caracterizează colectivitatea A (vezi capitolul X) ca, de pildă, media aritmetică, mediana, dispersia, modulul etc. În raport cu obiectivul pe care-l urmărim ne interesează mărimea uneia din aceste valori tipice. Distribuția exactă a colectivității A nu o cunoaștem, astfel că nu sîntem în măsură să o calculăm. De aceea, recurgem la metoda selecției, extrăgînd la întîmplare un sondaj x, format din n unități.

Colectivitatea x astfel extrasă are drept variante ale caracteristicii, valorile $x_1, x_2, \ldots, x_n$.

Fie a valoarea medie a colectivității A :

$$a = \frac{a_1 + a_2 + \ldots + a_N}{N} = \frac{1}{N} \sum_{i=1}^{N} a_i . \qquad \text{(XX, 2.1)}$$

Cu ajutorul colectivității extrase x ne propunem să obținem o apreciere a lui a cît mai exactă.

Pentru aceasta considerăm valoarea medie a valorilor lui x :

$$\bar{x} = \frac{x_1 + x_2 + \ldots + x_n}{n} . \qquad \text{(XX, 2.2)}$$

Media de selecție $\bar{x}$ este o variabilă aleatoare în sensul calculului probabilităților, deoarece variază de la sondaj la sondaj.

În statistica matematică se demonstrează următoarea teoremă: *Într-un sondaj simplu, de o singură caracteristică, avem relațiile* :

$$M(\bar{x}) = a, \qquad \text{(XX, 2.3)}$$

$$D^2(\bar{x}) = (1 - f)\frac{s^2}{n} \qquad \text{(XX, 2.4)}$$

unde

$$s^2 = \frac{1}{N-1} \sum_{i=1}^{N} (a_i - a)^2, \qquad f = \frac{n}{N} . \qquad \text{(XX, 2.5)}$$

Demonstrația teoremei se bazează pe schema probabilistică a bilei nerevenite — selecția *nerepetată*; ea poate fi dată pornindu-se de la distribuția hipergeometrică (vezi Cap. VII).

Se știe că

$$\left.\begin{aligned} M(\bar{x}) &= a, \\ D^2(\bar{x}) &= \frac{N-n}{N-1} \cdot \frac{\sigma^2}{n}. \end{aligned}\right\} \qquad (XX, 2.6)$$

Eliminînd pe

$$\sum_{i=1}^{N}(a_i - a)^2$$

între relația

$$\sigma^2 = \frac{1}{N}\sum_{i=1}^{N}(a_i - a)$$

și relația (XX, 2.5), obținem

$$\sigma^2 = \frac{N-1}{N} s^2 .$$

Înlocuind pe σ^2, relațiile (XX, 2.6) devin:

$$M(\bar{x}) = a$$

$$D^2(\bar{x}) = (1 - f)\frac{s^2}{n} .$$

Se vede că pentru n suficient de mare

$$D^2(\bar{x}) \to 0 .$$

Teorema este demonstrată, în cazul particular cînd valorile a_i sînt egale cu 1 sau 0. În mod analog se face demonstrația și în cazul general.

Această teoremă, fundamentală pentru teoria selecției, ne spune lucrul următor:

a) Valoarea medie a lui $\bar{x}$ este tocmai valoarea medie a, a colectivității generale.

b) Dispersia lui $\bar{x}$ dată prin egalitățile (XX, 2.4) și (XX, 2.5) este în general un număr mic, deoarece conține la numitor numărul n, care în general este un număr mare, iar s este de ordinul de mărime al valorilor a_i.

Legătura dintre $\bar{x}$ și a devine acum în baza inegalității clasice a lui Cebîșev, cunoscută.

Probabilitatea ca $\bar{x}$ să fie cuprins între $a - k\sqrt{D^2(\bar{x})}$ și $a + k\sqrt{D^2(\bar{x})}$ este mai mare decît $1 - \frac{1}{k^2}$. Cum $\sqrt{D^2(\bar{x})}$, în baza relației (XX, 2.4), este foarte mic, înseamnă că $\bar{x}$ oscilează într-un interval foarte restrîns în jurul valorii a și că acest interval este cu atît mai mic cu cît numărul unităților din sondaj este mai mare.

Scanned with OKEN Scanner

Se justifică astfel teoretic motivul pentru care poate fi folosită metoda selecției, deoarece se demonstrează că valoarea obținută $\bar{x}$ poate fi întrebuințată ca o valoare aproximativă a lui a, media caracteristicii studiate, în colectivitatea generală.

*Precizia**) *probabilistică a sondajului simplu nerepetat.* Ne propunem spre exemplu, să determinăm venitul mediu anual al familiilor dintr-o comună. Fie $N = 2\,000$ numărul familiilor din această comună. Pentru cele 2 000 familii să admitem că venitul mediu anual este $a = 12\,000$ lei. Valoarea lui a am obținut-o din formula (XX, 2.1) făcînd media aritmetică a tuturor veniturilor anuale a_i, $(i = 1, 2, \ldots, N)$.

Să efectuăm acum o selecție de 5% din numărul familiilor colectivității totale. Vom alege deci, la întîmplare, 100 familii din întreaga colectivitate de 2 000 familii. Calculînd media aritmetică a veniturilor anuale ale acestor 100 familii să presupunem că am găsit, prin aplicarea formulei (XX, 2.2), $\bar{x} = 11\,580$.

Se înțelege de la sine că dacă vom lua o altă selecție de 100 familii vom găsi, în general pentru $\bar{x}$, o altă valoare, de pildă, $\bar{x} = 12\,750$. Deci, $\bar{x}$ variază de la sondaj la sondaj și se pune problema cu ce precizie putem aproxima noi valoarea $a = 12\,000$ cu ajutorul unei valori $\bar{x}$, rezultată din aplicarea unui singur sondaj.

Formulele (XX, 2.3) și (XX, 2.4) ne dau media și dispersia lui $\bar{x}$. Se vede însă din formulele (XX, 2.3), (XX, 2.4), (XX, 2.5), (XX. 2.6) că noi nu putem calcula cu elementele pe care le avem la dispoziție pe $M(\bar{x})$ și $D^2(\bar{x})$. Într-adevăr, $M(\bar{x})$ și $D^2(\bar{x})$, în baza formulelor respective, se calculează cu ajutorul caracteristicilor a_i, $(i = 1, 2, \ldots, N)$, ale întregii colectivități, pe care nu le cunoaștem.

Pentru noi, în baza sondajului de mărime n pe care l-am făcut, sînt cunoscute numai caracteristicile x_i, $(i = 1, 2, \ldots, n)$, și numai în baza acestor valori trebuie să determinăm precizia selecției.

Valoarea exactă a lui $D^2(\bar{x})$, prin urmare, n-o cunoaștem. Aceasta nu înseamnă însă că nu putem să determinăm o valoare aproximativă a ei cu un ordin de aproximație cunoscut.

Să notăm prin $\bar{s}^2$ expresia

$$\bar{s}^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2. \qquad \text{(XX, 2.7)}$$

Întrucît $\bar{s}^2$ se exprimă cu ajutorul valorilor x_i, $(i = 1, 2, \ldots, n)$, care variază de la sondaj la sondaj, $\bar{s}^2$ este de asemenea o variabilă aleatoare, a cărei valoare medie și dispersie nu prezintă dificultăți de calcul.

Într-adevăr, se poate arăta că asimptotic

$$M(\bar{s}^2) \simeq s^2 \qquad \text{(XX, 2.8)}$$

*) Aici ca și mai departe noțiunea de „precizie" a sondajului se referă numai la gradul de mărime a erorilor de reprezentativitate întîmplătoare adică la variația acestor erori în limita unui interval stabilit cu o anumită probabilitate.

unde s^2 este dat de formula (XX, 2.5) și că

$$D^2(\bar{s}^2) \cong \frac{c^2}{n}, \qquad \text{(XX, 2.9)}$$

unde c este o constantă, exprimată în raport cu valorile a_i, $(i = 1, 2, \ldots, n)$.

Vom demonstra aceste rezultate pentru expresia mai generală

$$\bar{m}_k = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^k}{n-1}$$

care pentru $k = 2$ coincide cu $\bar{s}^2$, în afară de factorul constant $\frac{n}{n-1}$.

Pentru simplificarea calculelor vom presupune că sondajul este făcut dintr-o populație teoretică infinită. Lăsăm în seama cititorului să aducă în calcule modificările corespunzătoare pentru cazul unei populații finite.

Notăm prin

$$\bar{M}_k = \frac{\sum_{i=1}^{n} x_i^k}{n}$$

și prin M_k valoarea medie a variabilei x_i^k :

$$M_k = M\left(x_i^k\right).$$

Rezultă imediat

$$M(\bar{M}_k) = M\left(\frac{\sum_{i=1}^{n} x_i^k}{n}\right) = \frac{\sum_{i=1}^{n} M\left(x_i^k\right)}{n} = M_k$$

$$M(\bar{M} - M_k)^2 = \frac{1}{n^2} M\left[\sum_{i=1}^{n} (x_i^k - M_k)^2\right] = \frac{1}{n^2} \sum_{i=1}^{n} M(x_i^k - M_k)^2 = \frac{M_{2k} - M_k^2}{n}.$$

Să observăm că

$$(x_i - \bar{x})^k = x_i^k - C_k^1 \bar{x} x_i^{k-1} + C_k^2 \bar{x}^2 x_i^{k-2} - \ldots$$

Rezultă că

$$\bar{m}_k = \bar{M}_k - C_k^1 \bar{x} M_{k-1} + C_k^2 \bar{x}^2 \bar{M}_{k-2} - \ldots$$

Să observăm că putem considera $M_k = m_k$ pentru orice k, deoarece putem presupune $M = 0$ fără a restrînge generalitatea. Pe de altă parte,

$$M(\bar{x}^s \bar{M}_{k-s}) \leqslant \sqrt{M(\bar{x}^{2s}) M(\bar{M}_{k-s}^2)}$$

conform inegalității lui Schwarz.

Însă :

$$M(\bar{M}_{k-s}^2) = M_{k-s}^2, \qquad M(\bar{x}^{2s}) = 0\left(\frac{1}{n^s}\right).$$

Prin urmare, expresia $M(\bar{x}^s \overline{M}_{k-s})$ este asimptotică cel mult de ordinul lui $\frac{1}{n^s}$ pentru orice s.

Din expresia lui $\overline{m}_k$ rezultă, atunci

$$M(\overline{m}_k) = m_k + 0\left(\frac{1}{n}\right).$$

Făcînd $k = 2$, găsim

$$M\left(\frac{n}{n-1}\,\bar{s}^2\right) = \frac{n}{n-1}\,s^2 + 0\left(\frac{1}{n}\right) = s^2 + 0\left(\frac{1}{n}\right).$$

Pe de altă parte,

$$D^2(\overline{m}_k) = M(\overline{m}_k - m_k)^2 = M(\overline{M}_k - m_k - k\bar{x}\overline{M}_{k-1} + C_k^2\bar{x}^2 M_{k-2} - \ldots)^2 =$$

$$= M(\overline{M}_k - m_k)^2 + k^2 M(\bar{x}^2\overline{M}_{k-1}^2) - 2kM(\bar{x}\overline{M}_{k-1}(\overline{M}_k - m_k)) + T,$$

unde s-a notat cu T restul termenilor care provin din ridicarea la pătrat a expresiei $(m_k - m_k)$.

Însă,

$$M(\overline{M}_k - m_k)^2 = \frac{m_{2k} - m_k^2}{n},$$

$$M(\bar{x}^2\overline{M}^2{}_{k-1}) = \frac{1}{n^4}\,M\left(\sum x_i + \sum x_i^{k-1}\right)^2 =$$

$$= \frac{1}{n^4}\,M\left(\sum_{i=1}^{n} x_i^2 + \sum_{i\neq j} x_i x_j\right)\left(\sum_{h=1}^{n} x_h^{2k-2} + \sum_{h\neq l}^{n} x_h^{k-1} x_l^{k-1}\right)$$

$$= \frac{1}{n^4}\,M\left[\sum_{i\neq h} x_i^2 x_h^{2k-2} + \sum_{i\neq h\neq l} x_i^2 x_h^{k-1} x_l^{k-1} + \sum_{i\neq j\neq h}^{n} x_i x_j x_h^{2k-2} \sum_{i\neq j\neq h\neq l}^{n} x_i x_j x_h^{k-1} x_l^{k-1}\right]$$

Întrucît $M(x_i) = 0$ și variabilele x_i sînt independente, termenii

$$\sum_{\substack{i,j,h=1\\ i\neq j\neq h}}^{n} x_i x_j x_h^{2k-2}, \qquad \sum_{\substack{i,j,h,k=1\\ i\neq j\neq h\neq l}}^{n} x_i x_j x_h^{k-1} x_l^{k-1}$$

dau valori medii nule.

Pe de altă parte,

$$M\left(x_i^2 x_h^{2k-2}\right) = m_2 m_{2k-2}, \qquad M\left(x_i^2 x_h^{k-1} x_l^{k-1}\right) = m_2 m_{k-1}^2,$$

pentru orice i, h, l diferiți între ei.

Rămîne să numărăm cîți termeni de forma $x_i^2 x_h^{2k-2}$ și cîți de forma $x_i^2 x_h^{k-1} x_l^{k-1}$, cînd indicii i, h, l iau în condițiile de mai sus, toate valorile de la 1 la n. Numărul termenilor de forma

$$x_i^2 x_h^{2k-2}, \qquad x_i^2 x_h^{k-1} x_l^{k-1}$$

este respectiv egal cu

$$n(n-1), \quad n(n-1)(n-2)$$

Prin urmare, valoarea asimptotică a momentului $M\left(\bar{x}^2 \overline{M}_{k-1}^2\right)$ va fi dată numai de grupul de termeni

$$\sum_{i \neq h \neq l} x_i^2 x_h^{k-1} x_l^{k-1},$$

pentru care

$$\frac{1}{n^4} M\left(\sum_{i \neq h \neq l} x_i^2 x_h^{k-1} x_l^{k-1}\right) = \frac{1}{n} m_2 m_{k-1}^2 + 0\left(\frac{1}{n^2}\right).$$

$$M(\bar{x}\overline{M}_{k-1}(\overline{M}_k - m_k)) = M(\bar{x}\overline{M}_{k-1}\overline{M}_k) - m_k M(\bar{x}\overline{M}_{k-1}) =$$

$$= \frac{1}{n^3} M\left(\sum_{i=1}^{n} x_i \sum_{i=1}^{n} x_i^{k-1} \sum_{i=1}^{n} x_i^k\right) - \frac{m_k}{n^2} M\left(\sum_{i=1}^{n} x_i \sum_{i=1}^{n} x_i^{k-1}\right).$$

Efectuînd calculele, vom observa că pentru evaluarea asimptotică a momentului unui produs de forma

$$M\left(\sum_{i=1}^{n} x_i \sum_{i=1}^{n} x_i^{k-1} \sum_{i=1}^{n} x_i^k\right)$$

va trebui să păstrăm termenii $x_i^{\alpha_i}$, $x_j^{\alpha_j}, \ldots$, în care intră cei mai mulți indici $i, j, \ldots$, cu excepția valorilor $\alpha_i, \alpha_j, \ldots$, egale cu unu, care conduc la un moment nul.

Rezultă că

$$\frac{1}{n^3} M\left(\sum_{i=1}^{n} x_i \sum_{i=1}^{n} x_i^{k-1} \sum_{i=1}^{n} x_i^k\right) = \frac{1}{n^3} M\left(\sum_{i \neq j} x_i^{k-1} x_j^{k+1} + \sum_{i \neq j} x_i^k x_j^k\right)$$

și

$$M\left(x_i^{k-1} x_j^{k+1}\right) = m_{k-1} m_{k+1}; \quad M\left(x_i^k x_j^k\right) = m_k^2.$$

Avem atunci $n(n-1)$ termeni $x_i^{k-1} x_j^{k+1}$ și $x_i^k x_j^k$. Deci,

$$\frac{1}{n^2} M\left(\sum_{i=1}^{n} x_i \sum_{i=1}^{n} x_i^{k-1} \sum_{i=1}^{n} x_i^k\right) = \frac{m_{k-1} m_{k+1} + m_k^2}{n} + 0\left(\frac{1}{n^2}\right).$$

De asemenea,

$$\frac{m_k}{n^2} M\left(\sum_{i=1}^{n} x_i \sum_{i=1}^{n} x_i^{k-1}\right) = \frac{m_k}{n^2} M\left(\sum_{i=1}^{n} x_i^k\right) = \frac{m_k^2}{n}.$$

Prin urmare, obținem

$$M(\bar{x}\overline{M}_{k-1}(\overline{M}_k - m_k)) = \frac{m_{k-1} m_{k+1}}{n} + 0\left(\frac{1}{n}\right).$$

A mai rămas să evaluăm valoarea asimptotică a momentelor date de T. Procedînd în același mod, se vede că

$$M(\bar{x}^2 \overline{M}_{k-2}(\overline{M}_k - m_k)) = 0\left(\frac{1}{n^2}\right).$$

Toți ceilalți termeni, care sînt cuprinși în T, vor avea valori medii de forma

$$\frac{1}{n^r} M\left(x_i^\alpha x_j^\beta x_h^\gamma x_l^\delta\right),$$

dacă $r > 5$ și

$$\frac{1}{n^r} M\left(x_i^\alpha x_j^\beta x_h^\gamma\right)$$

dacă $r = 5$.

Se vede că asimptotic acești termeni sînt cel puțin de ordinul lui $\frac{1}{n^2}$.

În sfîrșit, ținînd seama de valorile medii pe care le-am calculat, rezultă că

$$D^2(\bar{m}_k) = M(\bar{m}_k - m_k)^2 = \frac{m_{2k} - m_k^2 - 2km_{k-1}m_{k+1} + k^2 m_2 m_{k-1}^2}{n} + 0\left(\frac{1}{n^2}\right).$$

Dacă facem $k = 2$, obținem un rezultat de forma

$$D^2\left(\frac{n}{n-1}\bar{s}^2\right) = \frac{c^2}{n} + 0\left(\frac{1}{n^2}\right).$$

De aci rezultă formula (XX,2.9).

În baza teoremei lui Cebîșev, rezultă că avem

$$s^2 - 3\frac{c}{\sqrt{n}} < \bar{s}^2 < s^2 + 3\frac{c}{\sqrt{n}}$$

cu o probabilitate mai mare decît $\frac{8}{9} = 0{,}89$.

Inegalitățile de mai sus se mai pot scrie:

$$\bar{s}^2 - 3\frac{c}{\sqrt{n}} < s^2 < \bar{s}^2 + 3\frac{c}{\sqrt{n}}. \qquad \text{(XX, 2.10)}$$

Pe de altă parte, tot din teorema lui Cebîșev rezultă în baza egalităților (XX, 2.3) și (XX, 2.4)

$$\bar{x} - 3\frac{s\sqrt{1-f}}{\sqrt{n}} < a < \bar{x} + 3\frac{s\sqrt{1-f}}{\sqrt{n}}$$

tot cu o probabilitate mai mare decît 0,89.

Dacă în relațiile de mai sus ținem seamă de (XX, 2.10), rezultă că putem scrie

$$\bar{x} - 3\frac{\bar{s}\sqrt{1-f}}{\sqrt{n}} < a < \bar{x} + 3\frac{\bar{s}\sqrt{1-f}}{\sqrt{n}}. \qquad \text{(XX, 2.11)}$$

În această inegalitate am lăsat la o parte termenii care asimptotic sînt mai mici decît ordinul $\frac{1}{\sqrt{n}}$.

Putem deci conchide că inegalitatea (XX, 2.11) este adevărată, lăsînd la o parte termenii mai mici decît ordinul $\frac{1}{\sqrt{n}}$, cu o probabilitate mai mare decît 0,89.

De aici rezultă precizia sondajului.

În exemplul pe care l-am ales, referitor la determinarea venitului mediu al familiilor dintr-o comună, să admitem că efectuînd calculele cuprinse în formula (XX, 2.7), găsim $\bar{s} = 1\,800$. Avem deci:

$$\bar{x} = 11\,580, \quad \bar{s} = 1\,800$$

$$n = 100, \; N = 2\,000, \; f = 0{,}05, \; 1-f = 0{,}975.$$

Aplicînd inegalitățile (XX, 2.11), obținem

$$11\,580 - 527 < a < 11\,580 + 527$$

sau

$$11\,053 < a < 12\,107.$$

Am estimat astfel venitul mediu al unei familii. El se găsește în intervalul (11 053—12 107 lei) cu o probabilitate mai mare de 0,89. Se observă că valoarea adevărată (12 000 lei) este în interiorul acestui interval.

2°. *Cazul cînd sînt studiate două caracteristici.* Colectivitatea generală A poate avea mai multe caracteristici.

Să considerăm astfel cazul a două caracteristici.

Colectivitatea generală are N unități și fiecare unitate are două caracteristici a_i și b_i, $(i = 1, 2, \ldots, N)$.

Referitor la această colectivitate vom avea:

$$a = \frac{a_1 + a_2 + \ldots + a_N}{N} \qquad \text{(XX, 2.12)}$$

$$b = \frac{b_1 + b_2 + \ldots + b_N}{N} \qquad \text{(XX, 2.13)}$$

$$\sigma_a^2 = \frac{1}{N}\sum_{i=1}^{N}(a_i - a)^2; \qquad s_a^2 = \frac{1}{N-1}\sum_{i=1}^{N}(a_i - a)^2 \qquad \text{(XX, 2.14)}$$

$$\sigma_{ab} = \frac{1}{N}\sum_{i=1}^{N}(a_i - a)(b_i - b); s_{ab} = \frac{1}{N-1}\sum_{i=1}^{N}(a_i - a)(b_i - b) \qquad \text{(XX, 2.15)}$$

$$\sigma_b^2 = \frac{1}{N}\sum_{i=1}^{N}(b_i - b)^2; \qquad s_b^2 = \frac{1}{N-1}\sum_{i=1}^{N}(b_i - b)^2. \qquad \text{(XX, 2.16)}$$

Formulele (XX, 2.12) și (XX, 2.13) ne dau respectiv mediile aritmetice ale caracteristicilor a_i și b_i, $(i = 1, 2, \ldots, N)$.

Formulele (XX, 2.14) și (XX, 2.16) reprezintă respectiv dispersiile caracteristicilor a_i și b_i, $(i = 1, 2, \ldots, N)$. Formula (XX, 2.15) servește la stabilirea corelației dintre cele două caracteristici.

Scanned with OKEN Scanner

Ne propunem să determinăm aproximativ valorile a și b pe baza unui sondaj simplu, cu n unități, care ne conduce la valorile $x_1, x_2, \ldots, x_n$, obținute la întîmplare. De data aceasta, fiecare dintre valorile x_i admite două caracteristici u_i și v_i $(i = 1,2,\ldots, n)$ și în funcție de ele ne propunem să calculăm pe a și b.

Notăm

$$\bar{u} = \frac{u_1 + u_2 + \ldots + u_n}{n},$$

$$\bar{v} = \frac{v_1 + v_2 + \ldots + v_n}{n};$$

$\bar{u}$ și $\bar{v}$ variază de la sondaj la sondaj și sînt, deci, niște variabile aleatoare.

De exemplu, să presupunem că A reprezintă colectivitatea familiilor dintr-o anumită regiune și că prima caracteristică a_i ne dă suma lunară cheltuită de familie pentru hrană, iar a doua caracteristică b_i ne dă venitul lunar al aceleiași familii. Pentru a evalua aceste valori pe întreaga regiune, folosim o selecție, bazată pe anchetarea unui număr mai redus $n < N$ de familii. Rezultatele obținute pe baza acestei selecții pot fi extinse pentru întreaga colectivitate? Există motive teoretice pentru a dovedi justețea folosirii metodei selecției în acest caz?

Răspunsul ni-l dă următoarea:

TEOREMĂ. *Într-un sondaj simplu, cu două caracteristici, avem relațiile:*

$$M(\bar{u}) = a, \qquad (XX, 2.17)$$

$$M(\bar{v}) = b, \qquad (XX, 2.18)$$

$$M(\bar{u} - a)^2 = (1 - f)\frac{s_a^2}{n}, \qquad (XX, 2.19)$$

$$M(\bar{v} - b)^2 = (1 - f)\frac{s_b^2}{n}, \qquad (XX, 2.20)$$

$$M(\bar{u} - a)(\bar{v} - b) = (1 - f)\frac{s_{ab}}{n}, \qquad (XX, 2.21)$$

unde $f = \frac{n}{N}$ $(n < N)$.

Primele patru relații se demonstrează în același mod ca în cazul sondajelor simple de unități cu o singură caracteristică. Rămîne să demonstrăm ultima relație.

Avem:

$$(\bar{u} - a)(\bar{v} - b) = \frac{1}{n^2}\sum_{i,j=1}^{n}(u_i - a)(v_j - b).$$

De aici deducem

$$M(\bar{u}-a)(\bar{v}-b)=\frac{1}{n^2}\sum_{i,j=1}^{n}M(u_i-a)(v_j-b)=\frac{1}{n^2}\sum_{i=1}^{n}M(u_i-a)(v_i-b)+$$

$$+\frac{1}{n^2}\sum_{i,j=1}^{n}M(u_i-a)(v_j-b)=\frac{1}{n}\sum_{h=1}^{N}(a_h-a)(b_h-b)P(x_i=l_h)+$$

$$+\frac{n-1}{n}\sum_{\substack{h,k=1\\h\neq k}}^{n}(a_h-a)(b_k-b)P(x_i=l_h,\ x_j=l_k),$$

unde prin l_j, $(j=1,2,\ldots,N)$, notăm vectorul

$$l_j=(a_j,\ b_j), \qquad (j=1,2,\ldots,N)$$

și prin $x_1, x_2, \ldots, x_n$ variabilele selecției

$$x_1=(u_1,\ v_1),\ x_2=(u_2,\ v_2),\ldots,x_n=(u_n,\ v_n)$$

avînd în vedere că, potrivit definiției lor,

$$M(u_i-a)(v_i-b)=\sum_{h=1}^{N}(a_h-a)(b_h-b)P(x_i=l_h)$$

și

$$M(u_i-a)(v_j-b)=\sum_{\substack{h,k=1\\h\neq k}}(a_h-a)(b_k-b)P(x_i=l_h,\ x_j=l_k).$$

Trebuie, deci, să calculăm și probabilitățile care intervin în formulele de mai sus. Astfel, probabilitatea ca u_i să ia valoarea a_h și v_i să ia valoarea b_h, adică probabilitatea ca vectorul $x_i(u_i, v_i)$ să ia valoarea vectorială $l_h(a_h, b_h)$ este

$$P(x_i=l_h)=\frac{1}{N}.$$

admițînd că $a_i\neq a_j$, pentru $i\neq j$.

În același mod avem

$$P(x_i=l_h,\ x_j=l_k)=\frac{1}{N(N-1)}, \qquad i\neq j,\ h\neq k.$$

Probabilitățile

$$P(x_i=l_h,\ x_i=l_k)$$
$$h\neq k$$

și

$$P(x_i=l_h,\ x_j=l_h)$$
$$i\neq j$$

sînt egale cu zero, deoarece prima reprezintă probabilitatea ca vectorul $x_i(u_i, v_i)$ să ia, în același timp, valorile vectoriale $l_h(a_h, b_h)$ și $l_k(a_k, b_k)$, ceea ce este o imposibilitate. A doua probabilitate reprezintă probabilitatea ca vectorii diferiți $x_i(u_i, v_i)$ și $x_j(u_j, v_j)$ să ia aceeași valoare vectorială $l_h(a_h, b_h)$, ceea ce de asemenea este imposibil, cunoscînd că sondajul se face fără a restitui unitatea extrasă.

Scanned with OKEN Scanner

Cu acestea, vom avea în continuare

$$M(\bar{u}-a)(\bar{v}-b)=\frac{1}{nN}\sum_{h=1}^{N}(a_h-a)(b_h-b)+$$

$$+\frac{n-1}{nN(N-1)}\sum_{\substack{h,k=1\\h\neq k}}^{N}(a_h-a)(b_k-b)=\frac{1}{nN}\sum_{h=1}^{N}(a_h-a)(b_h-b)+$$

$$+\frac{n-1}{nN(N-1)}\sum_{h=1}^{N}(a_h-a)\sum_{\substack{k=1\\k\neq h}}^{N}(b_k-b).$$

Însă, din egalitatea evidentă

$$\sum_{k=1}^{N}(b_k-b)=0,$$

deducem

$$\sum_{\substack{k=1\\h\neq k}}^{N}(b_k-b)=-(b_h-b).$$

În sfîrșit, obținem

$$M(\bar{u}-a)(\bar{v}-b)=\frac{1}{nN}\sum_{h=1}^{N}(a_h-a)(b_h-b)-\frac{n-1}{nN(N-1)}\sum_{h=1}^{N}(a_h-a)(b_h-b)=$$

$$=\frac{N-n}{N}\cdot\frac{1}{n(N-1)}\sum_{h=1}^{N}(a_h-a)(b_h-b)=(1-f)\frac{s_{ab}}{n}$$

relație pe care ne-am propus să o demonstrăm.

Se vede că pentru estimarea parametrului necunoscut a este suficient să ne bazăm pe egalitățile (XX, 2.17) și (XX, 2.19). Din ele rezultă că valoarea medie a lui $\bar{u}$ este a, iar dispersia lui $\bar{u}$ este mică datorită numitorului n. Deci, aplicîndu-se inegalitatea lui Cebîșev, rezultă că $\bar{u}$ se găsește, cu o probabilitate mare, într-un interval mic în jurul lui a.

La fel se estimează și $\bar{v}$.

În multe cazuri ne interesează raportul dintre caracteristicile a și b. De pildă, în exemplul considerat mai sus, prezintă un interes evident raportul dintre suma lunară cheltuită de o familie pentru hrană și venitul lunar al familiei. Raportul acesta pentru întreaga colectivitate este $\frac{a}{b}$ și se pune în mod firesc întrebarea dacă avem dreptul să admitem că $\frac{\bar{u}}{\bar{v}}$ reprezintă o valoare aproximativă a acestui raport. Răspunsul la această întrebare ni-l dă următoarea:

– TEOREMĂ. *Într-un sondaj simplu, ale cărui unități prezintă două caracteristici, avem egalitățile asimptotice*

$$M\left(\frac{\bar{u}}{\bar{v}}\right)\simeq\frac{a}{b}, \qquad \text{(XX, 2.22)}$$

$$D^2\left(\frac{\bar{u}}{\bar{v}}\right)\simeq(1-f)\frac{a^2}{nb^2}\left(\frac{s_a^2}{a^2}-2\frac{s_{ab}}{ab}+\frac{s_b^2}{b^2}\right). \qquad \text{(XX, 2.23)}$$

Presupunînd existența derivatelor parțiale de primele două ordine în raport cu $\bar{u}$ și $\bar{v}$ ale funcției $f(\bar{u}, \bar{v})$, putem aplica dezvoltarea după formula lui Taylor în jurul punctului $(\bar{u} = a, \bar{v} = b)$:

$$f(\bar{u}, \bar{v}) = f(a, b) + (\bar{u} - a) f'_{\bar{u}}(a, b) + (\bar{v} - b) f'_{\bar{v}}(a, b) +$$

$$+ \frac{1}{2} [(\bar{u} - a)^2 f''_{\bar{u}\bar{u}}(\xi, \eta) + 2(\bar{u} - a)(\bar{v} - b) f''_{\bar{u}\bar{v}}(\xi, \eta) + (\bar{v} - b)^2 f''_{\bar{v}\bar{v}}(\xi, \eta)],$$

(ξ, η) fiind un punct intermediar între punctele (a, b) și $(\bar{u}, \bar{v})$. În cazul de față, $f(\bar{u}, \bar{v}) = \frac{\bar{u}}{\bar{v}}$ și formula precedentă devine, presupunînd $b \neq 0$,

$$\frac{\bar{u}}{\bar{v}} = \frac{a}{b} + \frac{1}{b}(\bar{u} - a) - \frac{a}{b^2}(\bar{v} - b) - \frac{1}{\eta^2}(\bar{u} - a)(\bar{v} - b) + \frac{\xi}{\eta^3}(\bar{v} - b)^2.$$

Dacă notăm cu R' și R'' respectiv mulțimea punctelor din spațiul selecției R_n, pentru care sînt îndeplinite inegalitățile

$$|\bar{u} - a| < \varepsilon, \qquad |\bar{v} - b| < \varepsilon,$$

unde ε este un număr pozitiv oricît de mic, aceste inegalități vor defini în spațiul R_n o regiune

$$\overline{R} = R' \cap R''$$

Considerînd în spațiul R_n punctul $M_{j_1 j_2 \ldots j_n}$, fiecare dintre indicii $j_1, j_2, \ldots, j_n$ luînd valori de la 1 la N, să determinăm probabilitatea ca acest punct să aparțină regiunii $\overline{R}$. Această probabilitate, în baza inegalității lui Boole, este

$$P(M_{j_1 j_2 \ldots j_n} \in \overline{R}) \geqslant P(M_{j_1 j_2 \ldots j_n} \in R') + P(M_{j_1 j_2 \ldots j_n} \in R'') - 1.$$

Observăm că, pentru determinarea probabilității $P(M_{j_1 j_2 \ldots j_n} \in \overline{R})$, trebuie să cunoaștem probabilitățile $P(M_{j_1 j_2 \ldots j_n} \in R')$ și $P(M_{j_1 j_2 \ldots j_n} \in R'')$.

După inegalitatea lui Cebîșev și avînd în vedere relațiile (XX, 2.19) avem

$$P(M_{j_1 j_2 \ldots j_n} \in R') = P(|\bar{u} - a| < \varepsilon) \geqslant 1 - \frac{M(\bar{u} - a)^2}{\varepsilon^2} = 1 - \frac{(1 - f) S_a^2}{n\varepsilon^2} = 1 - \frac{h}{n},$$

unde

$$h = \frac{(1 - f) s_a^2}{\varepsilon^2}.$$

În mod analog,

$$P(M_{j_1 j_2 \ldots j_n} \in R'') \geqslant \left(1 - \frac{k}{n}\right),$$

unde

$$k = \frac{(1 - f) s_b^2}{\varepsilon^2}$$

de unde urmează că

$$P(M_{j_1 j_2 \ldots j_n} \in \overline{R}) \geqslant 1 - \frac{l}{n}$$

dacă $l = h + k$ și, evident,

$$P(M_{j_1 j_2 \ldots j_n} \in C\overline{R}) < \frac{l}{n}$$

$C\overline{R}$ fiind complementara regiunii $\overline{R}$.

Însemnînd, în general, prin

$$M_{\bar{R}}\,[F(\bar{u},\,\bar{v})]$$

fracțiunea din valoarea medie a unei funcții $F(\bar{u},\,\bar{v})$ pe regiunea $\bar{R}$ din spațiul R_n, vom scrie, în particular pentru dezvoltarea

$$M_{\bar{R}}\left(\frac{\bar{u}}{\bar{v}}\right) = M_{\bar{R}}\left(\frac{a}{b}\right) + \frac{1}{b}\,M_{\bar{R}}(\bar{u}-a) - \frac{a}{b^2}\,M_{\bar{R}}(\bar{v}-b) -$$

$$- M_{\bar{R}}\left[\frac{1}{\eta^2}(\bar{u}-a)(\bar{v}-b)\right] + M_{\bar{R}}\left[\frac{\xi}{\eta^3}(\bar{v}-b)^2\right].$$

Urmează, deci, să calculăm fiecare dintre fracțiunile valorilor medii pe regiunea $\bar{R}$, care intră în această dezvoltare, atît în membrul întîi cît și în membrul al doilea.

În primul rînd să observăm, conform ipotezei noastre ($a_i \geqslant 0,\ b_j \geqslant 0\,;\ i,j = 1,2,\ldots,N$), că mediile $\bar{u}$ și $\bar{v}$ sînt mărginite superior, adică satisfac inegalitățile

$$\bar{u} < k',\quad \bar{v} < k''.$$

În al doilea rînd, observăm, în baza aceleiași ipoteze, că raportul $\dfrac{\bar{u}}{\bar{v}}$ este de asemenea mărginit superior în tot spațiul R_n și deci, implicit, în subspațiul $\bar{R}\in R_n$, deoarece avem

$$\frac{\bar{u}}{\bar{v}} < \frac{k'}{b_m} = k''',$$

unde b_m reprezintă cea mai mică valoare dintre toate valorile $b_j > 0$, $(j = 1,2,\ldots,N)$ iar k''' este bineînțeles o constantă.

În sfîrșit, mai observăm că funcțiile $\dfrac{1}{\eta^2}$ și $\dfrac{\xi}{\eta^3}$ sînt continue în vecinătatea punctului $(\xi = a,\ \eta = b)$, întrucît din ipoteza de mai sus deducem

$$a = \frac{1}{N}\sum_{i=1}^{N} a_i \geqslant 0, \qquad b = \frac{1}{N}\sum_{j=1}^{N} b_j > 0.$$

Atunci, în virtutea continuității, putem alege un ε atît de mic, încît avînd

$$|\,\bar{u}-a\,| < \varepsilon, \qquad |\,\bar{v}-b\,| < \varepsilon$$

să existe un număr finit L, care să mărginească superior funcțiile $\dfrac{1}{\eta^2}$ și $\dfrac{\xi}{\eta^3}$:

$$\frac{1}{\eta^2} < L, \qquad \frac{|\,\xi\,|}{|\,\eta^3\,|} < L.$$

În urma acestor considerații, fracțiunile din valorile medii ale dezvoltării $M_{\bar{R}}\left(\dfrac{\bar{u}}{\bar{v}}\right)$ se calculează ușor. Într-adevăr, din primul membru al acestei dezvoltări deducem

$$M_{\bar{R}}\left(\frac{\bar{u}}{\bar{v}}\right) = M\left(\frac{\bar{u}}{\bar{v}}\right) - M_{C\bar{R}}\left(\frac{\bar{u}}{v}\right).$$

Pe de altă parte,

$$\left|\,M_{C\bar{R}}\left(\frac{\bar{u}}{\bar{v}}\right)\,\right| \leqslant M_{C\bar{R}}\left(\left|\,\frac{\bar{u}}{\bar{v}}\,\right|\right) < k'''\,P\,|\,M_{j_1 j_2 \ldots j_n} \in C\bar{R}) < \frac{k'''l}{n},$$

ținînd seamă de inegalitățile stabilite pînă acum.

Rezultă, deci,

$$M_{\bar{R}}\left(\frac{\bar{u}}{\bar{v}}\right) = M\left(\frac{\bar{u}}{\bar{v}}\right) + O\left(\frac{1}{n}\right).$$

Procedînd la calcularea membrului al doilea al dezvoltării obținem,

$$M_{\bar{R}}\left(\frac{a}{b}\right) = \frac{a}{b} P(M_{j_1 j_2 \ldots j_n} \in \bar{R}) \geqslant \frac{a}{b}\left(1 - \frac{l}{n}\right)$$

sau

$$M_R\left(\frac{a}{b}\right) = \frac{a}{b} + O\left(\frac{1}{n}\right).$$

Din egalitățile

$$M_{\bar{R}}(\bar{u} - a) + M_{C\bar{R}}(\bar{u} - a) = M(\bar{u} - a) = 0$$

deducem

$$M_{\bar{R}}(\bar{u} - a) = - M_{C\bar{R}}(\bar{u} - a).$$

Cum, însă, din relațiile de mai sus avem

$$| M_{C\bar{R}}(\bar{u} - a) | \leqslant M_{C\bar{R}} | \bar{u} - a | < k' P(M_{j_1 j_2 \ldots j_n} \in C\bar{R}) < \frac{k'l}{n}$$

rezultă

$$\frac{1}{b} M_{\bar{R}}(\bar{u} - a) = O\left(\frac{1}{n}\right).$$

În același fel găsim

$$\frac{a}{b^2} M_{\bar{R}}(\bar{v} - b) = O\left(\frac{1}{n}\right).$$

Ultimele două fracții ale valorilor medii din membrul al doilea al dezvoltării expresiei $M_{\bar{R}}\left(\frac{\bar{u}}{\bar{v}}\right)$ se calculează aplicînd mai întîi inegalitatea lui Schwarz

$$M_{\bar{R}}\left[\frac{1}{\eta^2}(\bar{u} - a)(\bar{v} - b)\right] \leqslant \sqrt{M_{\bar{R}}\left(\frac{1}{\eta^4}\right)} \cdot \sqrt{M_{\bar{R}}(\bar{u} - a)^2 (\bar{v} - b)^2}.$$

Deoarece avem

$$M_{\bar{R}}\left(\frac{1}{\eta^4}\right) \leqslant L^2 \cdot P(M_{j_1 j_2 \ldots j_n}) < L^2$$

și

$$M_{\bar{R}}(\bar{u} - a)^2 (\bar{v} - b)^2 = O\left(\frac{1}{n^2}\right)$$

rezultă

$$M_{\bar{R}}\left[\frac{1}{\eta^2}(\bar{u} - a)(\bar{v} - b)\right] = O\left(\frac{1}{n}\right).$$

Fără nici o deosebire, se demonstrează că

$$M_{\bar{R}}\left[\frac{\xi}{\eta^3}(\bar{v} - b)^2\right] = O\left(\frac{1}{n}\right).$$

După cele stabilite mai sus, dezvoltarea căutată devine

$$M\left(\frac{\bar{u}}{\bar{v}}\right) = \frac{a}{b} + O\left(\frac{1}{n}\right).$$

Scanned with OKEN Scanner

Am demonstrat astfel relația (XX, 2.22).

Avem, mai departe

$$\frac{\bar{u}}{\bar{v}} - \frac{a}{b} = \left[\frac{1}{b}(\bar{u}-a) - \frac{a}{b^2}(\bar{v}-b)\right] - \left[\frac{1}{\eta^2}(\bar{u}-a)(\bar{v}-b) - \frac{\xi}{\eta^3}(\bar{v}-b)^2\right].$$

Ridicînd la pătrat, obținem

$$\left(\frac{\bar{u}}{\bar{v}} - \frac{a}{b}\right)^2 = \frac{1}{b^2}(\bar{u}-a)^2 - \frac{2a}{b^3}(\bar{u}-a)(\bar{v}-b) + \frac{a^2}{b^4}(\bar{v}-b)^2 + R,$$

unde

$$R = -2\left[\frac{1}{b}(\bar{u}-a) - \frac{a}{b^2}(\bar{v}-b)\right]\left[\frac{1}{\eta^2}(\bar{u}-a)(\bar{v}-b) - \frac{\xi}{\eta^3}(\bar{v}-b)^2\right] + \left[\frac{1}{\eta^2}(\bar{u}-a)(\bar{v}-b) - \frac{\xi}{\eta^3}(\bar{v}-b)^2\right]^2.$$

Dispersia raportului $r = \frac{\bar{u}}{\bar{v}}$ va fi deci

$$D^2\left(\frac{\bar{u}}{\bar{v}}\right) = \frac{1}{b^2}M(\bar{u}-a)^2 - \frac{2a}{b^3}M(\bar{u}-a)(\bar{v}-b) + \frac{a^2}{b^4}M(\bar{v}-b)^2 + M(R) =$$

$$= \frac{1}{b^2}(1-f)\frac{s_a^2}{n} - \frac{2a}{b^3}(1-f)\frac{s_{ab}}{n} + \frac{a^2}{b^4}(1-f)\frac{s_b^2}{n} + M(R) =$$

$$= (1-f)\frac{a^2}{b^2 n}\left(\frac{s_a^2}{a^2} - 2\frac{s_{ab}}{ab} + \frac{s_b^2}{b^2}\right) + M(R)$$

substituind mediile $M(\bar{u}-a)^2$, $M(\bar{u}-a)(\bar{v}-b)$ și $M(\bar{v}-b)^2$ prin valorile lor date de relațiile (XX, 2,19), (XX, 2. 20), (XX,2. 21).

Calculele făcute după același procedeu ca mai sus și pe care nu le mai reproducem aici ne arată că

$$M(R) = O\left(n^{-\frac{3}{2}}\right)$$

și prin urmare dispersia raportului $r = \frac{\bar{u}}{\bar{v}}$ are ca expresie

$$D^2\left(\frac{\bar{u}}{\bar{v}}\right) = (1-f)\frac{a^2}{b^2 n}\left(\frac{s_a^2}{a^2} - 2\frac{s_{ab}}{ab} + \frac{s_b^2}{b^2}\right) + O\left(n^{-\frac{3}{2}}\right)$$

adică relația (XX, 2.23).

Egalitățile de mai sus sînt aproximative, în sensul că nu au loc decît pentru valorile mari ale lui n. Tot cu ajutorul inegalității lui Cebișev, ele ne arată că $\frac{\bar{u}}{\bar{v}}$ variază cu o probabilitate mare în jurul lui $\frac{a}{b}$, într-un interval foarte mic.

În sfîrșit, este interesant de observat că coeficientul de corelație a variabilelor $\bar{u}$ și $\bar{v}$ este dat de relația :

$$\rho_{\bar{u},\bar{v}} = \frac{s_{ab}}{s_a s_b}$$

și deci este independent de n, adică de mărimea sondajului.

Determinarea preciziei selecției în cazul a două caracteristici. Să revenim din nou la exemplul pe care l-am folosit mai sus, presupunînd că cele două caracteristici pentru fiecare familie sînt:

a) suma lunară cheltuită pentru consumul de alimente al unei familii;

b) venitul lunar al unei familii.

Pe baza unei selecții statistice, luînd la întîmplare 5% din cele 2 000 de familii din comună, ne propunem să determinăm, de pildă, valoarea medie a raportului dintre suma lunară cheltuită pentru consumul de alimente al unei familii și venitul ei lunar, pentru întreaga colectivitate de familii din comună.

Să notăm prin $\bar{u}$ și $\bar{v}$, următoarele expresii:

$$\bar{u} = \frac{u_1 + u_2 + \ldots + u_n}{n},$$

$$\bar{v} = \frac{v_1 + v_2 + \ldots + v_n}{n}.$$

Observăm că pentru determinarea preciziei selecției, formulele (XX, 2.22) și (XX, 2.23) sînt inaplicabile din cauză că elementele din membrul al doilea al egalității (XX, 2.23) sînt calculate cu valorile a_i și b_i ale caracteristicilor pe care le poartă unitățile din colectivitatea generală, și pe care nu le cunoaștem.

Din această cauză considerăm formulele:

$$\bar{s}_a^2 = \frac{1}{n-1} \sum_{i=1}^{n} (u_i - \bar{u})^2, \qquad (XX, 2.24)$$

$$\bar{s}_{ab} = \frac{1}{n-1} \sum_{i=1}^{n} (u_i - \bar{u})(v_i - \bar{v}), \qquad (XX, 2.25)$$

$$\bar{s}_b^2 = \frac{1}{n-1} \sum_{i=1}^{n} (v_i - \bar{v})^2. \qquad (XX, 2.26)$$

Repetînd atunci același raționament care ne-a condus la relația (XX, 2.11) putem afirma că avem:

$$\frac{\bar{u}}{\bar{v}} - 3\frac{\bar{u}}{\bar{v}}\sqrt{\frac{1-f}{n}} \cdot \sqrt{\frac{\bar{s}_1^2}{\bar{u}^2} - 2\frac{\bar{s}_{ab}}{\bar{u}\bar{v}} + \frac{\bar{s}_b^2}{\bar{v}^2}} < \frac{a}{b} < \frac{\bar{u}}{\bar{v}} +$$

$$+ 3\frac{\bar{u}}{\bar{v}}\sqrt{\frac{1-f}{n}} \cdot \sqrt{\frac{\bar{s}_a^2}{\bar{u}^2} - 2\frac{\bar{s}_{ab}}{\bar{u}\bar{v}} + \frac{\bar{s}_b^2}{\bar{v}^2}}, \qquad (XX, 2.27)$$

cu o probabilitate mai mare decît 0,89.

Formula (XX, 2.27) ne dă precizia sondajului cu privire la evaluarea raportului $\frac{a}{b}$.

Dacă ne referim numai la evaluarea caracteristicilor medii a și b, avem în baza relației (XX, 2.11):

$$\bar{u} - 3\frac{\bar{s}_a\sqrt{1-f}}{\sqrt{n}} < a < \bar{u} + 3\frac{\bar{s}_a\sqrt{1-f}}{\sqrt{n}} \qquad \text{(XX, 2.28)}$$

$$\bar{v} - 3\frac{\bar{s}_b\sqrt{1-f}}{\sqrt{n}} < b < \bar{v} + 3\frac{\bar{s}_b\sqrt{1-f}}{\sqrt{n}} \qquad \text{(XX, 2.29)}$$

fiecare în parte, fiind verificată cu o probabilitate mai mare de 0,89.

Formulele (XX, 2.28) și (XX, 2.29) sînt utile pentru că arată precizia sondajului privind caracteristicile a și b.

O altă problemă pe care ne propunem s-o rezolvăm este următoarea: cunoscînd precizia evaluării caracteristicilor a și b să se determine precizia raportului $\frac{a}{b}$ și în general a unei funcții date $f(a, b)$.

Cu alte cuvinte, cunoscînd relațiile (XX, 2.28) și (XX, 2.29) să determinăm un interval probabil de variație, care să cuprindă raportul $\frac{a}{b}$.

Pentru rezolvarea acestei probleme vom folosi o metodă nouă, care prezintă anumite avantaje:

Să notăm:

$$\delta = 3\cdot\frac{\bar{s}_a\sqrt{1-f}}{\sqrt{n}},$$

$$\varepsilon = 3\cdot\frac{\bar{s}_b\sqrt{1-f}}{\sqrt{n}}.$$

Inegalitățile (XX, 2.28) și (XX, 2.29) se pot scrie:

$$\bar{u} - \delta < a < \bar{u} + \delta, \qquad \text{(XX, 2.28')}$$

$$\bar{v} + \varepsilon < b < \bar{v} + \varepsilon. \qquad \text{(XX, 2.29')}$$

Deducem:

$$\frac{\bar{u}-\delta}{\bar{v}+\varepsilon} < \frac{a}{b} < \frac{\bar{u}+\delta}{\bar{v}-\varepsilon}. \qquad \text{(XX, 2.30)}$$

Însă:

$$\frac{\bar{u}+\delta}{v-\varepsilon} = \frac{\bar{u}+\delta}{\bar{v}}\cdot\frac{1}{1-\frac{\varepsilon}{\bar{v}}} = \frac{\bar{u}+\delta}{\bar{v}}\left(1+\frac{\varepsilon}{\bar{v}}+\ldots\right) = \frac{\bar{u}}{\bar{v}} + \frac{\delta}{\bar{v}} + \frac{\bar{u}\varepsilon}{\bar{v}^2} + \ldots$$

$$\frac{\bar{u}-\delta}{\bar{v}+\varepsilon} = \frac{\bar{u}-\delta}{\bar{v}}\cdot\frac{1}{1+\frac{\varepsilon}{\bar{v}}} = \frac{\bar{u}-\delta}{\bar{v}}\left(1-\frac{\varepsilon}{\bar{v}}+\ldots\right) = \frac{\bar{u}}{\bar{v}} - \frac{\delta}{\bar{v}} - \frac{\bar{u}\varepsilon}{\bar{v}^2} + \ldots$$

Scanned with OKEN Scanner

În aceste formule s-au lăsat deoparte termenii care conțin pe δ și ε la o putere mai mare decît 1. Ținînd seama de ele, relația (XX, 2.30) devine

$$\frac{\bar{u}}{\bar{v}} - \frac{\delta}{\bar{v}} - \frac{\bar{u}\varepsilon}{\bar{v}^2} < \frac{a}{b} < \frac{\bar{u}}{\bar{v}} + \frac{\delta}{\bar{v}} + \frac{\bar{u}\varepsilon}{\bar{v}^2}. \qquad (XX, 2.31)$$

Se poate demonstra că precizia dată de relația (XX, 2.27) este întotdeauna mai mare decît precizia dată de (XX, 2.31).

Într-adevăr, pentru aceasta trebuie să avem

$$\left|\frac{\delta}{\bar{v}} + \frac{\bar{u}\varepsilon}{\bar{v}^2}\right| \geqslant \left|3\frac{\bar{u}}{\bar{v}}\sqrt{\frac{1-f}{n}} \cdot \sqrt{\frac{\bar{s}_a^2}{\bar{u}^2} - \frac{2\bar{s}_{ab}^2}{\bar{u}\bar{v}} + \frac{\bar{s}_b^2}{\bar{v}^2}}\right|$$

sau, ținînd seama de valorile lui δ și ε, după ce facem simplificările

$$\left|\frac{\bar{s}_a}{\bar{v}} + \frac{\bar{u}}{\bar{v}^2}\bar{s}_b\right| \geqslant \left|\frac{\bar{u}}{\bar{v}}\sqrt{\frac{\bar{s}_a^2}{\bar{u}^2} - 2\frac{\bar{s}_{ab}}{\bar{v}\bar{u}} + \frac{\bar{s}_b^2}{\bar{v}^2}}\right|.$$

Inegalitatea de mai sus va avea loc în același timp și sens cu ambii membri ridicați la pătrat. Efectuînd calculele, rezultă astfel că trebuie să avem:

$$\frac{2\bar{u}}{\bar{v}^2}(\bar{s}_a\bar{s}_b + \bar{s}_{ab}) \geqslant 0.$$

Dacă notăm

$$\rho_{ab} = \frac{\bar{s}_{ab}}{\bar{s}_a\bar{s}_b} \qquad (XX, 2.32)$$

trebuie deci să avem

$$\frac{2\bar{u}}{\bar{v}^3} \cdot \bar{s}_a\bar{s}_b(1 + \bar{\rho}_{ab}) \geqslant 0. \qquad (XX, 2.33)$$

Inegalitatea de mai sus este întotdeauna adevărată. Într-adevăr, numerele $\bar{u}$, $\bar{v}$, $\bar{s}_a$, $\bar{s}_b$ sînt pozitive. Pe de altă parte, $\bar{\rho}_{ab}$ este coeficientul de corelație al valorilor $\bar{u}_i$, $(i = 1, 2, \ldots, n)$ și $\bar{v}_i$, $(i = 1, 2, \ldots, n)$. După cum se știe, valoarea coeficientului de corelație este întotdeauna cuprinsă între -1 și 1, astfel că avem $1 + \bar{\rho}_{ab} \geqslant 0$.

Ne interesează cazul cînd diferența de precizie dată de (XX, 2.33) este cît mai mică.

Considerînd că $\bar{u} > 0$, $\bar{v} > 0$, observăm că inegalitatea (XX, 2.33) se transformă în egalitate dacă:

$$\bar{\rho}_{ab} = -1.$$

Deci, dacă coeficientul de corelație este egal cu -1, prin metoda pe care o propunem, ajungem la aceeași precizie a sondajului ca și prin metoda directă, dată de formula (XX, 2.27). Se știe că dacă coeficientul de corelație este egal cu -1, între variabile, la care el se referă, există cea mai

strînsă relație de dependență inversă. Anume, dacă una din variabile crește în valoare, cealaltă descrește proporțional.

Prin urmare, pentru caracteristicile care admit o puternică dependență inversă, avînd coeficientul de corelație -1, precizia dată prin metoda propusă nu diferă esențial de precizia obținută prin aplicarea metodei generale (XX, 2.27).

Pe de altă parte, primul membru din relația (XX, 2.33) este în general cu mult mai mică decît 1, după cum se poate vedea ușor. Într-adevăr ne interesează de obicei valorile pentru care

$$\frac{\bar{u}}{\bar{v}} < 1.$$

Pe de altă parte,

$$\frac{\bar{s}_b}{\bar{v}} < 1$$

deoarece reprezintă coeficientul de variație a lui $\bar{v}$. La fel vom avea

$$\frac{\bar{s}_a}{\bar{v}} < 1,$$

deoarece $\bar{v} > \bar{u}$.

Produsul acestor trei cantități sensibil mai mici decît 1 ne dă un număr mult mai mic decît 1, de unde rezultă că diferența dintre preciziile obținute prin ambele metode este în general mică.

Se pune în mod natural întrebarea cu ce probabilitate au loc inegalitățile (XX, 2.31). Vom observa că aceste inegalități sînt îndeplinite odată ce sînt îndeplinite inegalitățile (XX, 2.28) și (XX, 2.29) care fiecare în parte se verifică cu o probabilitate mai mare decît 0,89.

Se știe, însă, în baza cunoscutei inegalități a lui Boole (vezi capitolul III) că dacă însemnăm prin p_1 probabilitatea ca să se întîmple un eveniment (α), prin p_2 probabilitatea ca să se întîmple un eveniment (β) și prin p_{12} probabilitatea ca să se întîmple (α) și (β) împreună, avem relația:

$$p_{12} \geqslant p_1 + p_2 - 1. \qquad \text{(XX, 2.34)}$$

Însemnînd atunci prin p_1 probabilitatea ca să se verifice inegalitatea (XX, 2.28), prin p_2 probabilitatea ca să se verifice inegalitatea (XX, 2.29) și prin p_{12} probabilitatea ca aceste inegalități să aibă loc simultan, avem

$$p_{12} \geqslant p_1 + p_2 - 1 > 0{,}89 + 0{,}89 - 1 = 0{,}78.$$

Rezultă, deci, că inegalitățile (XX, 2.31) au loc cu o probabilitate mai mare decît 0,78.

Metoda preconizată prezintă importanță în cazul cînd ni se cere să estimăm pe baza unui sondaj o funcție oarecare $f(a, b)$ de variabile independente a și b. Cazul pe care l-am considerat

$$f(a, b) = \frac{a}{b},$$

este simplu, deoarece el se găsește expus prin alte metode și în tratatele clasice de statistică matematică.

Avantajul metodei pe care o propunem este că, spre deosebire de metodele cunoscute, permite estimarea cu un anumit grad de precizie a unei funcții oarecare de a și b.

De pildă, ne propunem să evaluăm printr-un sondaj de mărime n, expresia

$$\frac{a-2b}{a+2b},$$

unde a și b sînt mediile caracteristicilor date (XX, 2.17) și (XX, 2.18).

Din relațiile (XX, 2.28) și (XX, 2.29) și corelațiile lor (XX, 2.28′) (XX, 2.29′)

$$\frac{\bar{u}-\delta-2(\bar{v}+\varepsilon)}{\bar{u}-\delta+2(\bar{v}+\varepsilon)}<\frac{a-2b}{a+2b}<\frac{\bar{u}+\delta-2(\bar{v}-\varepsilon)}{\bar{u}-\delta+2(\bar{v}-\varepsilon)},$$

sau limitîndu-se numai la puterea întîi a lui δ și ε:

$$\frac{\bar{u}-2\bar{v}}{\bar{u}+2\bar{v}}-\frac{2\bar{u}(\delta+2\varepsilon)}{(\bar{u}+2\bar{v})^2}<\frac{a-2b}{a+2b}<\frac{\bar{u}-2\bar{v}}{\bar{u}+2\bar{v}}+\frac{2\bar{u}(\delta+2\varepsilon)}{(\bar{u}+2\bar{v})^2}$$

și înlocuind pe δ și ε prin valorile lor:

$$\frac{\bar{u}-2\bar{v}}{\bar{u}+2\bar{v}}-\frac{6\,\bar{u}\sqrt{1-f}(\bar{s}_a+2\,\bar{s}_b)}{\bar{u}+2\,\bar{v})^2}<\frac{a-2b}{a+2b}<\frac{\bar{u}-2\bar{v}}{\bar{u}+2\bar{v}}+$$

$$+\frac{6\,\bar{u}\sqrt{1-f}\,(\bar{s}_a+2\bar{s}_b).}{(\bar{u}+2\,\bar{v})^2} \qquad \text{(XX, 2.35)}$$

Această inegalitate, după formula lui Boole, are loc, după cum am văzut mai sus, cu o probabilitate mai mare de 0,78.

Expresia $\frac{\bar{u}-2\bar{v}}{\bar{u}+2\bar{v}}$, calculată cu elementele sondajului, constituie o estimație a raportului $\frac{a-2b}{a+2b}$, după cum ne arată inegalitățile (XX, 2.35) care în același timp ne dau și precizia sondajului.

Nu este cazul să insistăm asupra faptului că metoda se aplică și cînd considerăm mai multe caracteristici decît două precum și în cazul cînd considerăm funcții mai generale decît cele de mai sus.

Am văzut că probabilitatea cu care au fost obținute rezultatele în cele două cazuri examinate (cazul cu o singură caracteristică și cu două caracteristici) am determinat-o pe baza teoremei lui Cebîșev care ne arată că rezultatele respective s-au obținut cu o probabilitate mai mare decît un număr dat.

Scanned with OKEN Scanner

Acestei probabilități i se poate da, după cum știm, o valoare mai precisă prin aplicarea funcției $\Phi(t)$.

$$\Phi(t) = \frac{1}{\sqrt{2\pi}} \int_{-t}^{+t} e^{-\frac{t^2}{2}} dt. \qquad (XX, 2.36)$$

Prin urmare, probabilitatea ca media generală a (ne referim la primul exemplu) să se încadreze între cele două limite fixate anticipat se obține prin relația:

$$P\left(\bar{x} - t\cdot\sqrt{\frac{s^2}{n}(1-f)} < a < \bar{x} + t\sqrt{\frac{s^2}{n}(1-f)}\right) \cong \frac{1}{\sqrt{2\pi}} \int_{-t}^{+t} e^{-\frac{t^2}{2}} dt \quad (XX, 2.37)$$

în care t este un coeficient oarecare reprezentînd raportul dintre eroarea limită și abaterea medie pătratică. Dacă notăm cu Δ_x eroarea limită, atunci,

$$t = \frac{\Delta_x}{\sqrt{D^2(\bar{x})}} = \frac{\Delta_x}{\sqrt{(1-f)\frac{s^2}{n}}}.$$

Probabilitatea este deci o funcție de t; valorile ei sînt date în tabelul anexă I. După cum știm, A. M. Liapunov a demonstrat că relația (XX, 2.37) se verifică și în cazuri mult mai generale, nu numai în cadrul schemei de selecție a urnei lui Bernoulli.

Relațiile de mai sus sînt valabile, bineînțeles, pentru valorile lui n destul de mari, pentru ca utilizarea distribuției normale să fie permisă. Mai trebuie să arătăm că în cazul cînd studiem un fenomen care se aseamănă cu urna clasică a lui Bernoulli cu două stări cu bila nerevenită — — cînd avem deci o *caracteristică alternativă* — formulele respective privind media de selecție, dispersia medie de selecție, abaterea medie pătratică a mediei de selecție (eroarea medie a mediei de selecție) și eroarea limită a mediei de selecție își schimbă forma în mod corespunzător.

Vom avea pentru medie și dispersia caracteristicii

$$M\left(\frac{w}{n}\right) = np; \qquad (XX, 2.38)$$

$$D^2\left(\frac{w}{n}\right) = (1-f)\frac{p(1-p)}{n}; \qquad (XX, 2.39)$$

iar pentru eroarea medie a mediei de selecție și pentru eroarea limită ce se poate admite în cazul cercetării selective;

$$\mu = \sqrt{D^2\left(\frac{w}{n}\right)} = \sqrt{(1-f)\cdot\frac{p(1-p)}{n}}, \qquad (XX, 2.40)$$

$$\Delta_{\frac{w}{n}} = t\cdot\sqrt{\left(1-f\right)\cdot\frac{p(1-p)}{n}}. \qquad (XX, 2.41)$$

3°. *Determinarea volumului sondajului.* Una din cele mai importante probleme ce se pune cu prilejul lucrărilor pregătitoare ale cercetării selective este stabilirea volumului selecției. La stabilirea lui statisticianul trebuie să țină seama de precizia pe care trebuie să o aibă rezultatele selecției, de probabilitatea cu care ele trebuie garantate, de urgența cu care se cere efectuată lucrarea ca și de cheltuielile pe care le ocazionează.

Este evident că precizia rezultatelor crește pe măsură ce mărim volumul selecției.

Într-adevăr, din relația

$$\Delta_x = t \cdot \sqrt{\frac{s^2}{n}(1-f)} \qquad (XX, 2.42)$$

deducem

$$\Delta_x^2 = t^2 \frac{s^2}{n} \cdot \left(1 - \frac{n}{N}\right) = \frac{t^2 s^2 - n \cdot \frac{t^2 s^2}{N}}{n}$$

de unde, în sfîrșit, avem

$$n = \frac{t^2 s^2}{\Delta_x^2 + \frac{t^2 s^2}{N}}. \qquad (XX, 2.43)$$

În mod analog se obține și formula corespunzătoare dacă se studiază o caracteristică alternativă

$$n = \frac{t^2 \cdot p(1-p)}{\Delta_w^2 + \frac{t^2 p(1-p)}{N}}. \qquad (XX, 2.44)$$

Din relația (XX, 2.42) nu este greu de observat că eroarea limită de selecție este invers proporțională cu pătratul volumului selecției. Aceasta înseamnă că dacă, de exemplu, vrem să dublăm precizia selecției, celelalte condiții rămînînd neschimbate, va trebui ca volumul n al selecției să-l mărim de $2^2 = 4$ ori.

3. Sondajul simplu repetat

Schema probabilistică a acestui sondaj este aceea a urnii Bernoulli cu bila revenită. Efectuînd un asemenea sondaj lucrăm în condiții mai puțin precise ca în cazul sondajului simplu nerepetat și, în general, ca în oricare alt fel de sondaj, întrucît în acest caz — după cum s-a mai arătat — făcîndu-se n extracții independente, fiecare unitate a colectivității generale poate fi extrasă și introdusă în selecție de mai multe ori, ceea ce, evident, are o influență directă asupra nivelului mediei de selecție, asupra dispersiei ei și asupra erorii de reprezentativitate întîmplătoare.

Teoremele fundamentale stabilite pentru sondajul simplu nerepetat sînt valabile, în general, și pentru acest sondaj.

Avem, de pildă, pentru cazul cînd unitățile sondajului prezintă o singură caracteristică:

$$M(\bar{x}) = a \qquad (XX, .3.1)$$

$$D^2(\bar{x}) = \frac{\sigma^2}{n} \qquad (XX, 3.2)$$

$$\overline{M}_2 = \sqrt{D^2(\bar{x})} = \sqrt{\frac{\sigma^2}{n}} = \frac{\sigma}{\sqrt{n}}. \qquad (XX, 3.3)$$

Demonstrația poate fi dată astfel:

Știm că media de selecție dată de expresia (XX, 3.1) este evident o funcție de variabilele $x_1, x_2, \ldots x_n$.

Să presupunem că $F(x)$ este funcția de repartiție a distribuției teoretice și aceasta din urmă are primele k momente finite. Fiecare dintre variabilele x_i, $(i = 1, 2, \ldots, n)$ corespunde unei distribuții identice cu distribuția teoretică. Ne propunem să determinăm momentele mediei de selecție $\bar{x}$, în funcție de momentele teoretice:

$$M_k = \int_{-\infty}^{+\infty} x^k \, dF(x).$$

Astfel vom avea

$$\overline{M}_1 = M(\bar{x}) = M\left(\frac{\sum_{i=1}^{n} x_i}{n}\right) = \frac{1}{n} \frac{\sum_{i=1}^{n} M(x_i)}{n} = M_1.$$

În același fel vom obține:

$$M^2 = M(\bar{x}^2) = M\left(\frac{\sum_{i=1}^{n} x_i}{n}\right)^2 = \frac{1}{n^2}\sum_{i=1}^{n} M(x_i^2) + \frac{1}{n^2}\sum_{i \neq j}^{n} M(x_i x_j) = \frac{M_2}{n} + \frac{n-1}{n} M_1^2.$$

Dispersia mediei de selecție rezultă imediat. Într-adevăr, putem scrie

$$\overline{\mu}^2 = D^2(\bar{x}) = \overline{M}_2 - M_1^2 = \frac{M_2}{n} + \frac{n-1}{n} M_1^2 - M_1^2 = \frac{M_2 - M_1^2}{n} = \frac{\sigma^2}{n}.$$

Observăm că lipsește factorul $(1-f)$ care se vede în formula (XX, 2.4), ceea ce arată că dispersia mediei de selecție în cazul sondajului repetat este mai mare.

În mod analog putem să determinăm și momentele centrate de ordin superior ale mediei de selecție $\bar{x}$. Valoarea exactă a momentului centrat de ordinul r al mediei de selecție $\bar{x}$ este

$$\overline{\mu}_{\bar{x}}^{r} = \sum \frac{r!}{j_2! j_3! \ldots j_r!} \cdot \frac{n_j}{n^r} \left(\frac{\mu_2}{2!}\right)^{j_2} \left(\frac{\mu_3}{3!}\right)^{j_3} \cdots \left(\frac{\mu_r}{r!}\right)^{j_r} \qquad (XX, 3.4)$$

suma fiind extinsă la toți indicii $j_2, j_3, \ldots, j_r$, întregi, pozitivi sau nuli, supuși condiției $2j_2 + 3j_3 + \ldots + rj_r = r$ și notînd:

$$j = j_2 + j_3 + \ldots + j_r$$

$$n_j = \begin{cases} n(n-1)\ldots(n-j+1), & (j = 1, 2, \ldots, n) \\ 0 & (j = n+1, n+2, \ldots) \end{cases}$$

În particular, primele momente ale mediei de selecție $\bar{x}$, se deduc din (XX, 3.4):

$$\mu_{\bar{x}}^3 = \frac{\mu_3}{n^2},$$

$$\mu_{\bar{x}}^4 = \frac{3\mu_2^2}{n^2} + \frac{\mu_4 - 3\mu_2^2}{n^3}.$$

Dacă abaterea medie pătratică a mediei de selecție dată de formula (XX, 2.9) o multiplicăm cu factorul t, obținem eroarea limită a mediei de selecție

$$\Delta_{\bar{x}} = \pm t \cdot \sqrt{\frac{\sigma^2}{n}}. \qquad \text{(XX, 3.5)}$$

În cazul cînd caracteristica studiată este *alternativă*, eroarea limită a mediei de selecție este

$$\Delta_{\frac{w}{n}} = \pm t \cdot \sqrt{\frac{p(1-p)}{n}}. \qquad \text{(XX, 3.6)}$$

Ca și la selecția nerepetată, problema determinării volumului selecției n este deosebit de importantă.

Acest volum se obține cu ajutorul formulelor:

$$n = \frac{t^2\sigma^2}{\Delta_{\bar{x}}^2}, \qquad \text{(XX, 3.7)}$$

cînd studiem o caracteristică nealternativă și

$$n = \frac{t^2 p(1-p)}{\Delta_{\frac{w}{n}}^2} \qquad \text{(XX, 3.8)}$$

cînd studiem o caracteristică alternativă.

4. Indicatorii erorilor relative ale sondajului

Cînd am prezentat metoda teoretică și practică de calcul al preciziei selecției, limitele intervalului de confidență au fost date în valori absolute, ale unităților de măsură, în care a fost exprimată și caracteristica studiată. Eroarea mediei de selecție poate fi exprimată însă și în mărimi relative.

De pildă, se poate introduce, în acest scop, coeficientul de variație care ne este cunoscut încă de la studiul distribuțiilor.

Astfel, coeficientul de variație a mediei de selecție poate fi exprimat prin formula:

$$V_{\bar{x}} = \frac{s_{\bar{x}}}{\bar{x}} \qquad \text{(XX, 4.1)}$$

în care $\bar{x}$, media de selecție, știm că reprezintă o estimație absolut corectă

a mediei din colectivitatea generală; în cazul unei caracteristici alternative, coeficientul de variație va fi exprimat prin formula

$$V_p = \frac{s_p}{p}, \qquad (XX, 4.2)$$

în care p este frecvența relativă de selecție care de asemenea estimează frecvența relativă a caracteristicii respective în colectivitatea generală.

În lucrările practice este deosebit de comodă folosirea pătratului coeficientului de variație căruia unii statisticieni îi spun *dispersia relativă*.

$$V_{\bar{x}}^2 = \frac{s_{\bar{x}}^2}{\bar{x}^2}. \qquad (XX, 4.3)$$

Folosind acest indicator pentru a estima dispersia mediei de selecție a unei caracteristici nealternative într-un sondaj simplu finit, vom scrie ținînd seama că $s_{\bar{x}}^2 = \frac{s^2}{n}(1-f)$:

$$V_{\bar{x}}^2 = \frac{s_{\bar{x}}^2}{\bar{x}^2} = \frac{\frac{s^2}{n}(1-f)}{\bar{x}^2} = \frac{s^2}{\bar{x}^2 n}(1-f) \qquad (XX, 4.4)$$

Să considerăm următorul exemplu: să presupunem că cercetăm selectiv venitul zilnic al unui colectiv de 500 muncitori în care scop facem un sondaj, alegînd la întîmplare 50 muncitori (10%). Înregistrăm următoarele date:

$$N = 500,\ n = 50,\ f = 0{,}10,\ 1 - f = 0{,}90$$

$$\bar{x} = 100 \text{ lei},\ s = 20 \text{ lei}.$$

și pe baza lor calculăm dispersia relativă a mediei de selecție:

$$V_{\bar{x}}^2 = \frac{s^2}{\bar{x}^2 n}(1-f) = \frac{20^2}{100^2 \cdot 50} \cdot 0{,}90 = 0{,}00072,$$

de unde

$$V_{\bar{x}} = \sqrt{0{,}0072} = 0{,}027.$$

După cum se vede, coeficientul de variație a venitului mediu zilnic al unui muncitor reprezintă 0,027 sau 2,7%. La folosirea acestui indicator al erorii relative putem de asemenea stabili intervalul de confidență în care se va găsi venitul mediu zilnic al unui muncitor în întreaga colectivitate de 500 muncitori.

Scanned with OKEN Scanner

Astfel, cu o probabilitate de 0,95, noi putem afirma că venitul mediu zilnic al unui muncitor în colectivitatea generală se va afla în intervalul :

$$100 \pm 1{,}96\,(0{,}027.100),$$

adică

$$100 + 5{,}29.$$

Cu alte cuvinte, venitul mediu zilnic al unui muncitor nu va fi mai mic de 94,71 lei și nici mai mare de 105,29 lei și acest lucru se afirmă cu o probabilitate de 95%.

Dispersia s^2 care intră în formula (XX, 4.1) și (XX, 4.2) este, evident, o funcție de selecție, folosită ca estimație aproximativă a dispersiei din colectivitatea generală σ^2. De aceea, precizia calculului lui V_x^2 va depinde de mărimea abaterii $s^2 - \sigma^2$. În unele cazuri vom avea abateri destul de sensibile care însă se pot înlătura prin mărirea volumului selecției. Același lucru se poate spune și în ceea ce privește V_p^2.

5. Cîteva observații generale asupra sondajului simplu

Sondajul simplu (repetat și nerepetat) este individual, deoarece pentru formarea lui se extrage numai cîte o singură unitate din colectivitatea generală, fără ca aceasta să fie împărțită în grupe tipice. Extracțiile se fac cu respectarea principiului selecției — adică cu totul la întîmplare iar caracterul întîmplător al selecției se asigură prin procedeul tragerii la sorți sau prin folosirea unor tabele de numere întîmplătoare, fiecare unitate avînd aceeași probabilitate de a fi aleasă în selecție.

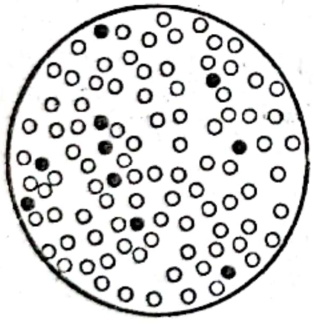


Fig. 42

Selecția întîmplătoare poate fi reprezentată grafic ca în figura 42 *).

Pentru teoremele fundamentale enunțate pentru cazul unor sondaje simple, s-a precizat și teorema că media de selecție

$$\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i$$

este o estimație *absolut corectă a mediei teoretice*

$$a = \frac{1}{N} \sum_{i=1}^{n} a_i .$$

*) Concepția acestui grafic ca și a altor grafice referitoare la diferite tipuri de selecții aparține autorilor sovietici : A. I. Boiarski, L. S. Brand, O. S. Davîdova, V. N. Starovski, B. I. Hotimski, B. S. Iastremski, vezi lucrarea *Statistica*, Soțekgiz, Moskva, 1936, p. 175 și urm.

Această teoremă răspunde principiului general al metodei selecției care constă în evaluarea caracteristicilor teoretice necunoscute ale colectivității generale cu ajutorul datelor rezultate dintr-un număr finit de experiențe. De obicei se consideră că valorile $x_1, x_2, \ldots, x_n$, reprezintă coordonatele unui punct $P(x_1, x_2, \ldots, x_n)$ dintr-un spațiu euclidian n-dimensional.

Fie $F(P)$ o funcție de selecție. Dacă

$$F(P) \to A$$
$$n \to \infty$$

convergența avînd loc *în probabilitate*, $F(P)$ reprezintă *o estimație* a valorii necunoscute A.

Faptul că convergența în medie pătratică implică convergența în probabilitate, prezintă o importanță deosebită în acest fel de probleme deoarece pentru a conchide că $F(P)$ este o estimație, este suficient să calculăm $M\,[F\,(P)]$ și $D^2\,[F\,(P)]$.

Datorită convergenței în probabilitate, cînd numărul observațiilor este mare, $F(P)$ este foarte apropiat de A, diferența dintre $F(P)$ și A fiind din ce în ce mai mică, atunci cînd $n \to \infty$. Valoarea lui $F(P)$ poate fi, deci, efectiv considerată ca o estimație a lui A. Există însă o infinitate de funcții $F(P)$ care tind în probabilitate către A. Vom spune că ele reprezintă estimații corecte ale lui A.

Dacă

$$M[F(P)] = A$$

și

$$D^2[F(P)] \to 0,$$

atunci $F(P)$ tinde în probabilitate către A.

Dacă

$$M[F(P)] = A + B(n)$$

și $B(n) \to 0$, cînd $n \to \infty$, convergența în probabilitate a lui $F(P)$ către A se păstrează cînd $n \to \infty$.

În primul caz, $F(P)$ este o *estimație absolut corectă* a lui A, iar în al doilea caz, o estimație corectă. Din cauza termenului $B(n)$ care tinde către 0 cînd $n \to \infty$, dar care pentru n finit, oricît de mare, are o valoare diferită de zero, estimația comportă o eroare sistematică.

Această eroare sistematică, dată de $B(n)$ este cu atît mai mică cu cît n este mai mare. În cazul estimației absolut corecte, eroarea sistematică dispare.

Teoria estimației reprezintă o cucerire recentă a științei în legătură cu metoda selecției. Am văzut că funcțiile de selecție au, în majoritatea cazurilor, proprietatea de a aproxima diferite valori caracteristice ale colectivității generale și că această aproximație este cu atît mai precisă, cu cît volumul selecției e mai mare. Definițiile pe care le-am dat pun o ordine în acest mod de aproximare, bazat pe convergența în probabilitate. Găsirea celor mai bune funcții de estimație pentru aceeași valoare caracteristică a populației totale constituie problema de bază a teoriei estimației.

Mai întîi ne putem ușor convinge că există mai multe funcții de estimație absolut corecte pentru același parametru. Ca exemplu luăm un eveniment de probabilitate constantă p în fiecare probă. Supunem evenimentul la n probe independente și cunoscînd rezultatele fiecărei probe, ne propunem pe baza lor să estimăm probabilitatea p, a cărei valoare numerică ne este necunoscută.

Scanned with OKEN Scanner

Împărțim grupul de n experiențe efectuate în două: în primul grup punem primele n_1 experiențe efectuate, iar în al doilea cele n_2 experiențe care ne-au mai rămas ($n_1 + n_2 = n$).

Fie ν_1 numărul de bile albe obținute în primele n_1 experiențe și ν_2 numărul de bile albe ieșite în cele n_2 experiențe ($\nu_1 + \nu_2 = \nu$).

Dacă k_1 și k_2 sînt două numere reale și pozitive date, alegem ca funcție de estimație a lui p expresia

$$\frac{k_1\nu_1 + k_2\nu_2}{k_1n_1 + k_2n_2}.$$

Efectuînd calculele, rezultă

$$M\left(\frac{k_1\nu_1 + k_2\nu_2}{k_1n_1 + k_2n_2}\right) = p, \qquad D^2\left(\frac{k_1\nu_1 + k_2\nu_2}{k_1n_1 + k_2n_2}\right) = \frac{k_1^2n_1 + k_2^2n_2}{(k_1n_1 + k_2n_2)^2}\,pq = 0\left(\frac{1}{n}\right).$$

Prin urmare,

$$\frac{k_1\nu_1 + k_2\nu_2}{k_1n_1 + k_2n_2}$$

reprezintă o funcție de estimație absolut corectă.

Observăm însă că

$$D^2\left(\frac{k_1\nu_1 + k_2\nu_2}{k_1n_1 + k_2n_2}\right) = D^2\left(\frac{\nu}{n}\right) + \frac{(k_1 - k_2)^2n_1n_2}{n(k_1n_1 + n_2k_2)^2}\,pq$$

deci

$$D^2\left(\frac{k_1\nu_1 + k_2\nu_2}{k_1n_1 + k_2n_2}\right) \geqslant D^2\left(\frac{\nu}{n}\right)$$

egalitatea avînd loc numai cînd $k_1 = k_2$ și $\frac{k_1\nu_1 + k_2\nu_2}{k_1n_1 + k_2n_2}$ devine egal cu $\frac{\nu}{n}$.

Toate funcțiile de estimație de forma

$$\frac{k_1\nu_1 + k_2\nu_2}{k_1n_1 + k_2n_2}$$

distincte de $\frac{\nu}{n}$ au dispersia mai mare decît $\frac{\nu}{n}$. Deși toate aceste funcții de estimație sînt absolut corecte, totuși cea preferabilă este aceea care dă naștere la cea mai mare concentrare a masei de probabilitate în jurul lui p, adică la cea mai mică dispersie. Prin urmare, vom alege pe $\frac{\nu}{n}$ ca funcție de estimație a lui p, în locul expresiilor mai sus considerate.

Am văzut că dispersia lui $\frac{\nu}{n}$ este mai mică decît dispersia funcției de estimație absolut corecte

$$\frac{k_1\nu_1 + k_2\nu_2}{k_1n_1 + k_2n_2}.$$

Nu am demonstrat însă că dispersia lui $\frac{\nu}{n}$ este mai mică decît dispersia oricărei alte funcții de estimație absolut corecte.

Dintre toate funcțiile de estimație absolut corecte, referitoare la același parametru, acea care dă naștere la dispersia cea mai mică poartă numele de *funcție de estimație eficientă*. Importanța acestei funcții rezidă în faptul că ea va da cea mai bună estimație a parametrului considerat. Teoria funcțiilor de estimație eficiente este dominată de teorema lui Rao-Cramer, care afirmă că dacă $f(x, \alpha)$ este o funcție de distribuție statistică continuă conținînd parametrul α și $\beta(x_1, x_2, \ldots, x_n)$ o funcție de selecție absolut corectă a lui α, există întotdeauna inegalitatea

$$D^2(\beta) \geqslant \frac{1}{n\int_{-\infty}^{+\infty}\left|\frac{\partial \lg f(x;\alpha)}{\partial \alpha}\right|^2 f(x,\alpha)\,dx}.$$

Funcțiile de estimație absolut corecte au deci, în virtutea teoremei lui Rao-Cramer, dispersia mărginită inferior. Funcțiile de estimație absolut corecte pentru care această limită este atinsă poartă numele de funcții de estimație eficiente.

În exemplul dat mai sus, funcția de estimație $\frac{\nu}{n}$ este într-adevăr eficientă.

Fie

$$\bar{\alpha}(\nu) = \frac{\nu}{n}.$$

Atunci, după cum am văzut,

$$M(\bar{\alpha}) = p,$$

$$D^2(\bar{\alpha}) = \frac{pq}{n}$$

valoarea $\frac{\nu}{n}$ este deci o funcție de estimație absolut corectă. Vom dovedi că această funcție de estimație este și eficientă.

Într-adevăr, din teorema lui Rao-Cramer știm că, oricare ar fi funcția de estimație corectă $\bar{\beta}(\nu)$, există inegalitatea

$$D^2(\bar{\beta}) \geqslant \frac{1}{n\sum_{i=1}^{n}\left(\frac{\partial \ln p_i}{\partial p}\right)^2 p_i(\alpha)}.$$

În cazul nostru parametrul α fiind egal cu p,

$$p_1(p) = p$$

$$p_2(p) = 1 - p$$

$$\sum_{i=1}^{n}\left(\frac{\partial \ln p_i}{\partial p_i}\right)^2 p_i = p\,\frac{1}{p^2} + (1-p)\,\frac{1}{(1-p)^2} = \frac{1}{pq}.$$

Prin urmare

$$D^2(\bar{\beta}) \geqslant \frac{pq}{n}.$$

Se vede că minimul este atins pentru $\frac{\nu}{n}$, care este o funcție de estimație eficientă.

★

Scanned with OKEN Scanner

Sondajul de serii. O variantă particulară a sondajului simplu este așa-numitul *sondaj de serii.* Sondajul simplu pe care l-am examinat pînă acum s-a format extrăgînd „la întîmplare" cîte o singură unitate din colectivitatea generală. Sînt cazuri, însă, cînd unitățile elementare ale colectivității generale formează ansambluri de obicei neomogene care se pot considera unități de observare *complexe.* În economia socialistă exemple de asemenea unități complexe sînt numeroase. De pildă, echipele de muncitori din exploatările miniere, brigăzile de producție și de intervenție de la întreprinderile de extracție a țițeiului, brigăzile de muncă din cadrul cooperativelor de producție agricole, grupe de populație după localități etc. pot fi privite, din punctul de vedere al cercetării statistice, ca unități de observare complexe. Cînd studiem selectiv colectivitățile formate din astfel de unități complexe, este evident că problema care se pune este aceea de a fi extrase cu respectarea principiului selecției, tocmai aceste unități complexe sau *serii* de unități individuale elementare. Efectuăm, astfel, un *sondaj de serii.*

O serie extrasă este asimilată în acest fel de sondaj cu unitatea simplă din celelalte feluri de sondaje; în cadrul sondajului, caracteristica studiată apare sub forma nivelului ei mediu din serie ($\bar{x}_i$) și de aceea, toate unitățile simple din care se compune seria extrasă sînt supuse înregistrării fără nici o excepție pentru ca nivelul mediu al caracteristicii în serie să fie cît mai real.

Dacă notăm cu r volumul selecției (adică un număr anumit de serii), atunci media caracteristicii în colectivitatea de selecție va fi

$$M(\bar{x}_i) = \frac{\sum_{i=1}^{r} \bar{x}_i}{r} \qquad \text{(XX, 5.1)}$$

iar dispersia mediei de selecție pe care o vom nota cu $\delta^2_{\bar{x}_i}$ va fi

$$D^2(\bar{x}_i) = \delta^2_{\bar{x}_i} = \frac{\sum_{i=1}^{r} [\bar{x}_i - M(x_i)]^2}{r} \qquad \text{(XX, 5.2)}$$

După cum se vede, prin forma ei, dispersia mediei de selecție de serie este identică cu cea de la sondajul simplu, prin conținutul ei, însă, ea este o dispersie *dintre grupe*, reflectînd numai variația caracteristicii care are loc sub influența schimbării factorilor obiectivi principali.

Din punct de vedere organizatoric sondajul de serii este mai convenabil și de aceea practicienii sînt înclinați a-i da preferință, față de celelalte metode de sondaj. Trebuie totuși avut în vedere că eroarea de selecție și volumul sondajului vor fi altele în cazul acestui sondaj, și de obicei, mai mari decît la sondajul individual.

Reprezentarea grafică a sondajului de serii egale este dată în figura 43.

Scanned with OKEN Scanner

Formulele de mai sus ale mediei și dispersiei de selecție în cazul unui sondaj de serii se referă la ipoteza cînd seriile sînt formate din numere egale de unități elementare.

Ne vom opri deci numai asupra unui sondaj de serii egale, avînd o schemă matematică mai simplă.

Eroarea medie de selecție pentru o caracteristică nealternativă va fi dată de formulele:

a) în cazul unui sondaj nerepetat

$$\mu = \sqrt{\frac{\delta_x^2}{r}\left(1 - \frac{r}{R}\right)} \qquad \text{(XX, 5.3)}$$

în care R reprezintă numărul unităților complexe (seriilor) care formează colectivitatea generală, iar celelalte simboluri ne sînt de-acum cunoscute;

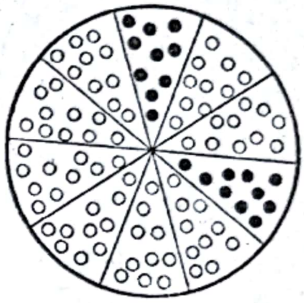


Fig. 43

b) în cazul unui sondaj repetat,

$$\mu = \sqrt{\frac{\delta_x^2}{r}}. \qquad \text{(XX, 5.4)}$$

Același indicator pentru o caracteristică alternativă va fi dat de formulele:

a) în cazul unui sondaj nerepetat

$$\mu = \sqrt{\frac{\delta_p^2}{r}\left(1 - \frac{r}{R}\right)}; \qquad \text{(XX, 5.5)}$$

b) în cazul unui sondaj repetat

$$\mu = \sqrt{\frac{\delta_p^2}{r}}. \qquad \text{(XX, 5.6)}$$

În aplicații practice este mai utilă folosirea coeficientului $\left(\frac{R-r}{R-1}\right)$ decît a celui dat în formulele de mai sus privind sondajul de serii nerepetat, întrucît are darul de a micșora — foarte puțin însă — eroarea întîmplătoare de reprezentativitate.

Sondajul de serii este cu atît mai precis cu cît, pe serii, caracteristica studiată are variații mai mici, în particular cînd variația pe serii este aceeași, sondajul devine absolut precis dar, în acest caz, organizarea lui nu mai prezintă nici un fel de avantaj.

În domeniul economiei unde lipsa de uniformitate în distribuția caracteristicilor este mare, sondajul de serii, de regulă, este mai puțin precis decît sondajul individual. El are, însă, avantajul de a putea fi mai ușor organizat.

În afară de sondajul de serii în forma arătată se mai poate folosi și o formă combinată a acestuia, în care, în prima fază, se alege un număr anumit de serii iar în faza a doua, din cadrul fiecărei serii, un anumit

număr de unități elementare, cu respectarea în ambele faze, a principiului selecției*).

Astfel, la stabilirea prin metoda sondajului a recoltei medii la hectar se poate folosi următorul procedeu : se alege — în prima fază — un anumit număr de sfaturi populare sau gospodării colective și abia în a doua fază, din cadrul acestora, se alege un procent determinat de sectoare care se supun cercetării.

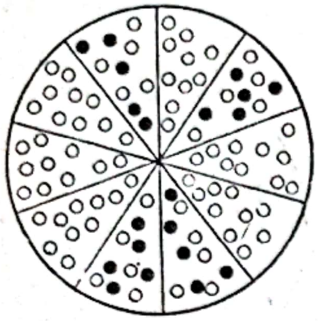


Fig. 44

Reprezentarea grafică a sondajului de serii egale, combinat, este dată în figura 44.

Eroarea medie de selecție în cazul sondajului de serii egale combinat este dată de formulele (demonstrația acestor formule este dată, într-un caz mai general, la capitolul privind sondajul în două faze) :

a) în cazul sondajului nerepetat

$$\mu = \sqrt{\frac{\bar{\sigma}^2}{n}\left(1 - \frac{n}{N}\right) + \frac{\bar{\delta}^2}{r}\left(1 - \frac{r}{R}\right)}; \qquad (XX, 5.7)$$

b) în cazul sondajului repetat

$$\mu = \sqrt{\frac{\bar{\sigma}^2}{n} + \frac{\bar{\delta}^2}{r}}, \qquad (XX, 5.8)$$

în care :

$\bar{\sigma}^2$ este mărimea medie a dispersiilor de serie;
$\bar{\delta}^2$ dispersia dintre serii, calculată după formula (XX, 5.2);
N numărul elementelor din fiecare serie,
n volumul selecției, iar
R și r ne sînt cunoscute.

6. Sondajul mecanic

Un alt fel de sondaj, simplu prin conținutul lui, este sondajul mecanic. El se aplică de obicei cînd unitățile colectivității generale sînt așezate într-o anumită ordine, de pildă, în ordinea crescîndă sau descrescîndă a variantelor caracteristicii studiate. O asemenea ordine poate fi alcătuită în mod conștient, prin înregistrarea în liste speciale a unităților colectivității totale **).

Împărțind volumul colectivității generale la volumul selecției (stabilit în funcție de precizia și probabilitatea cu care vrem să obținem rezultatele cercetării), obținem un număr k, numit *pasul de numărare* după care se extrag unitățile și se includ în sondaj.

Acest procedeu de selectare definește acel sondaj care în literatura de specialitate este numit *sondaj mecanic*.

*) A. I. Boiarski și alții, *op. cit.*, p. 191—192.

**) Acest lucru este posibil dacă colectivitatea generală a mai fost supusă cercetării în alt scop decît acela pe care-l urmărește cercetarea selectivă dată.

Caracterul selecției unităților în cazul sondajului mecanic, depinde de structura colectivității generale. Sondajul mecanic poate fi asimilat unui sondaj stratificat (tipic), în care din fiecare grupă tipică se alege cîte o singură unitate. El diferă însă de un sondaj stratificat prin faptul că grupele în care se împarte colectivitatea nu sînt tipice în adevăratul înțeles al cuvîntului deși, într-o oarecare măsură, ele sînt omogene. Dacă ordinea în care sînt așezate unitățile colectivității (crescîndă sau descrescîndă) este întîmplătoare, atunci toate tipurile, prin structura lor, vor fi asemănătoare, dacă se face abstracție de deosebiri cu totul întîmplătoare. În acest caz, prin esența lui, sondajul mecanic nu este decît un sondaj simplu și de aceea calculul erorilor de selecție se va face la fel ca în cazul acestuia din urmă. În situația inversă însă, cînd ordinea unităților nu este întîmplătoare, sondajul mecanic nu va fi pur întîmplător. În aceste condiții, sondajul conținînd cîte o unitate de fiecare tip, va fi mai preferabil sondajului simplu.

Eficacitatea sondajului mecanic, în sensul posibilității ce o poate oferi ca erorile de reprezentativitate să fie minime, depinde în mod hotărîtor de caracterul colectivității generale și de procedeul de selecții ales. De pildă, sondajul mecanic poate da rezultate în totul nesatisfăcătoare dacă elementele colectivității generale sînt afectate de o periodic itate oarecare Acest pericol poate deveni deosebit de grav dacă datele privind diferit elemente ale colectivității generale se înregistrează în ordine cronologică.

În cazul sondajului mecanic se aplică de obicei formulele de calcul stabilite pentru sondajul simplu repetat. Trebuie subliniat faptul că, folosind aceste formule, mărim precizia selecției, întrucît erorile date de sondajul mecanic sînt în general mai mici decît cele ale sondajului simplu repetat, însă ceva mai mari decît ale sondajului simplu nerepetat.

Un exemplu, devenit clasic, de aplicare a selecției mecanice este luarea probelor cu rama metrică dintr-un lan de grîu pentru a stabili, prin metoda sondajului, ce recoltă medie se va obține la recoltare. Nu vom insista mai mult asupra acestui fel de sondaj a cărei metodologie este descrisă amănunțit în diferite lucrări de specialitate *).

Reprezentarea grafică a selecției mecanice este dată în figura 45.

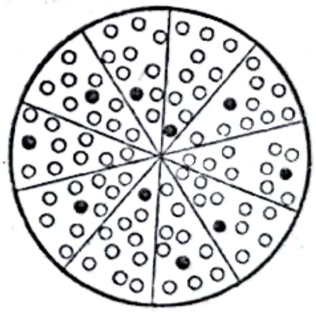


Fig. 45

7. Sondajul tipic (stratificat) simplu

În scopul obținerii unor rezultate mai precise, practica a impus folosirea și a unor alte sisteme de selecție. Printre aceste sisteme, o mare utilitate prezintă selecția tipică a cărei particularitate constă în faptul că, în prima etapă colectivitatea generală se desparte în mai multe *grupe tipice* numite și *straturi*, după una sau mai multe caracteristici, cu aplicarea științifică a principiilor

*) Vezi, de pildă, M. Biji și I. Stoichiță, *Metoda selectivă în cercetarea statistică*. Ed. științifică, București, 1957, p. 119–124.

grupării, iar în etapa a doua, din fiecare grupă astfel formată se efectuează cîte un sondaj care, de cele mai multe ori, nu este proporțional cu mărimea grupei.

Unitățile care formează fiecare *sondaj de grupă* se aleg individual, una cîte una, prin tragere la sorți sau cu aplicarea selecției mecanice.

Prin însumarea numărului de unități alese în fiecare grupă se obține *volumul sondajului general* pe baza căruia se stabilesc toți indicatorii sintetici ce ne sînt necesari pentru a caracteriza colectivitatea generală.

Este evident că procedeul expus, în care se îmbină principiul selecției cu principiile grupării tipice, asigură sondajului o reprezentativitate mult mai bună decît aceea cu care se lucrează în cazul unui sondaj simplu negrupat.

Vom considera aici numai sondajul tipic *nerepetat*, fiind cazul aproape general întîlnit în practică, în două ipoteze, după cum obiectul cercetării îl formează o singură caracteristică sau două caracteristici corelate.

1°. *Cazul cînd este studiată o singură caracteristică.* Din punct de vedere matematic, problema se pune în felul următor:

Să considerăm colectivitatea A împărțită în k subcolectivități parțiale:

$$A_1, A_2, \ldots, A_k.$$

Vom numi aceste subcolectivități *grupe* sau *straturi*.

În fiecare dintre cele k grupe corespund următoarele valori ale caracteristicii:

$$A_1(a_{11}, a_{21}, \ldots, a_{N_1 1})$$

$$A_2(a_{12}, a_{22}, \ldots, a_{N_2 2})$$

$$A_k(a_{1k}, a_{2k}, \ldots, a_{N_k k}).$$

După cum se vede, stratul A_1 are N_1 unități, A_2 are N_2 unități etc. Numărul total al unităților colectivității A este:

$$N_1 + N_2 + \ldots + N_k = N.$$

Din fiecare dintre aceste straturi facem, respectiv, cîte

$$n_1, n_2, \ldots, n_k$$

extracții la întîmplare, astfel ca

$$n_1 + n_2 + \ldots + n_k = n$$

fără ca unitatea extrasă să fie înapoiată în colectivitatea generală.

Prin urmare, din fiecare strat efectuăm cîte un sondaj obținînd, astfel, k sondaje ale căror unități au caracteristici cu valorile respective:

$$x_{11}, x_{21}, \ldots, x_{n_1 1},$$

$$x_{12}, x_{22}, \ldots, x_{n_2 2},$$

$$x_{1k}, x_{2k}, \ldots, x_{n_k k},$$

Scanned with OKEN Scanner

unde, evident, x_{ij}, $(i = 1, 2, \ldots, n_j\,;\ j = 1, 2, \ldots, k)$, sînt de considerat ca variabile aleatoare.

Introducînd notațiile :

$$a = \frac{1}{N} \sum_{j=1}^{k} \sum_{i=1}^{N_j} a_{ij}$$

$$a_j = \frac{1}{N_j} \sum_{i=1}^{N_j} a_{ij}$$

rezultă că

$$a = \frac{1}{N} \sum_{j=1}^{k} N_j a_j . \qquad \text{(XX, 7.1)}$$

Deci, media variantelor caracteristicii din colectivitatea generală poate fi interpretată ca media ponderată a mediilor de grupă ale caracteristicii, ponderile fiind respectiv egale cu

$$\frac{N_j}{N}, \qquad (j = 1, 2, \ldots, k).$$

În mod analog, notînd :

$$\bar{x} = \frac{1}{n} \sum_{j=1}^{k} \sum_{i=1}^{n_j} x_{ij}$$

$$\bar{x}_j = \frac{1}{n_j} \sum_{i=1}^{n_j} x_{ij}$$

deducem

$$\bar{x} = \frac{1}{n} \sum_{j=1}^{k} n_j \bar{x}_j ,$$

deci media valorilor caracteristicii din sondajul general (de mărimea $\sum_{j=1}^{k} n_j = n$), este egală cu media ponderată a mediilor de grupă ale valorilor caracteristicii din fiecare sondaj — ponderea fiind respectiv egală cu $\frac{n_j}{n}$, $(j = 1, 2, \ldots, k)$.

Revenind la cazul general, pentru evaluarea mediei generale a printr-un sondaj simplu tipic, vom considera variabila

$$y = \frac{1}{N} \sum_{j=1}^{k} N_j \bar{x}_j \qquad \text{(XX, 7.2)}$$

care reprezintă o medie ponderată a rezultatelor x_j, obținute în fiecare grupă. Valoarea y ne dă o evaluare a mediei a în baza teoremei de mai jos :

Scanned with OKEN Scanner

TEOREMĂ : *Într-un sondaj simplu, tipic avem*

$$M(y) = a \qquad \text{(XX, 7.3)}$$

$$D^2(y) = \sum_{j=1}^{k} \frac{N_j^2}{N^2}(1 - f_j)\frac{s_j^2}{n_j}, \qquad \text{(XX, 7.4)}$$

unde

$$s_j^2 = \frac{1}{N_j - 1}\sum_{j=1}^{N_j}(a_{ij} - a_j)^2; \quad \left(f_j = \frac{n_j}{N_j}\right). \qquad \text{(XX, 7.5)}$$

Din această teoremă rezultă condițiile de aplicare ale sondajului simplu tipic.

Teorema de mai sus care indică o estimație absolut corectă a mediei generale a poate fi ușor demonstrată.

Într-adevăr, în baza celor stabilite la studiul sondajului simplu, avem pentru grupa A_j relațiile :

$$M(\bar{x}_j) = a_j,$$

$$D^2(\bar{x}_j) = (1 - f_j)\frac{s_j^2}{n_j};$$

cunoscînd că variabilele $\bar{x}_j$ sînt independente între ele, deoarece din fiecare grupă extracțiile se efectuează independent unele de altele, putem scrie :

$$M(y) = \frac{1}{N}\sum_{j=1}^{k} N_j M(\bar{x}_j) = \frac{1}{N}\sum_{j=1}^{k} N_j a_j = a$$

și

$$D^2(y) = \frac{1}{N^2} D^2\left[\sum_{j=1}^{k} N_j(\bar{x}_j - a)\right]^2 = \frac{1}{N^2}\sum_{j=1}^{k} N_j^2 D^2(\bar{x}_j) = \sum_{j=1}^{k}\frac{N_j^2}{N^2}(1 - f_j)\frac{s_j^2}{n_j},$$

avînd în vedere relațiile (XX, 7.1) și (XX, 7.2).

După cum se vede, am stabilit relațiile (XX, 7.4) și (XX, 7.5) din teorema enunțată.

Pentru a putea aplica inegalitatea lui Cebîșev, trebuie ca dispersia lui y să fie un număr foarte mic. După cum rezultă însă din (XX, 7.4) acest lucru nu se întîmplă decît dacă numerele n_j sînt mari, iar dispersiile de grupă s_j^2 sînt mici. Prin urmare, ca sondajul simplu tipic să poată da rezultate acceptabile este necesar și suficient ca numărul unităților extrase din fiecare grupă să fie destul de mare.

Putem să observăm că dacă numerele n_j, $(j = 1, 2, \ldots, k)$, sînt suficient de mari, atunci $D^2(y)$ este foarte mică, și, prin urmare, numai în aceste condiții se poate afirma că variabila aleatoare y este o estimație *absolut corectă* a mediei valorilor caracteristicii din colectivitatea generală.

Această observație are o deosebită importanță practică.

Să considerăm următoarea aplicație :

Într-o întreprindere industrială s-a hotărît să se studieze, statistic, mărimea medie a întreruperilor în timpul de lucru al muncitorilor pe timp de o zi. În acest scop, s-a recurs la metoda selecției simple grupate.

Colectivitatea muncitorilor a fost împărțită în grupe tipice, pe meserii, s-au ales apoi cîte 50 de muncitori din fiecare grupă și s-a cronometrat timpul de lucru și de întrerupere în fiecare grupă selectată. Pe baza datelor obținute, se cere să se determine intervalul în care se va găsi mărimea medie a întreruperii în timpul de lucru zilnic la întreaga colectivitate de muncitori.

S-au obținut datele din tabelul 89 (primare și calculate) pe cele patru grupe de meserii.

Tabelul 89

Meseria	Numărul muncitorilor în fiecare grupă	Numărul muncitorilor cuprinși în selecție	Proporția selecției pe grupe	Numărul mediu de minute de întrerupere	Abaterea medie pătratică de grupă (minute)
k_i, $(i = 1,2,3,4)$	N_j	n_j	$\frac{n_j}{N_j} = f_j$	$\bar{x}_j$	s_j
	1	2	3	4	5
Strungari	500	50	0,10	30	8
Frezori	200	50	0,25	40	10
Ajustori	150	50	0,33	20	7
Rectificatori	350	50	0,14	30	9
Σ	$1\,200 = N$	$200 = n$	0,16	$30,41 = \bar{x}$	

Vom avea

$$M(\bar{x}) = \frac{1}{N}\sum_{i=1}^{k} N_j x_j = \frac{15\,000 + 8\,000 + 3\,000 + 10\,500}{1\,200} = 30,41.$$

Pentru estimarea mediei în colectivitatea generală ($\bar{X}$) vom considera eroarea limită a mediei de selecție cu siguranță probabilistică de 0,95, căruia îi corespunde $t = 1,96$.

În primul rînd, însă, trebuie calculată abaterea medie pătratică a mediei de selecție a sondajului general, care — după cum știm — va fi dată de expresia $\sqrt{D^2(\bar{x})}$, adică — ținînd seama de formulele (XX, 7.3) și (XX, 7.4) — vom avea:

$$\mu = \sqrt{\sum \frac{N_j^2}{N^2}(1 - f_j)\frac{s_j^2}{n_j}}. \qquad \text{(XX, 7.6)}$$

Pentru a determina valoarea lui μ în exemplul nostru ne vom servi și de date calculate din tabelul 90.

Introducînd în formula (XX, 7.6) datele concrete, obținem

$$\mu = \sqrt{0,363069} \simeq 0,60$$

iar pentru eroarea limită, cu probabilitatea 0,95,

$$\Delta_x = t\mu = 1,96 \cdot 0,60 = 1,176 \text{ minute} \simeq 1,18 \text{ minute}.$$

Scanned with OKEN Scanner

Tabelul 90

Meseria	$\frac{N_j}{N}$	$\left(\frac{N_j}{N}\right)^2$	$1-f_j$	s_j^2	$\frac{s_j^2}{n_j}$	$\frac{N_j^2}{N^2}(1-f_j)\frac{s_j^2}{n_j}$
a	1	2	3	4	5	6
Strungari	0,41	0,168	0,90	64	1,28	0,193536
Frezori	0,17	0,028	0,75	100	2,00	0,042000
Ajustori	0,13	0,016	0,67	49	0,98	0,010505
Rectificatori	0,29	0,084	0,86	81	1,62	0,117028
Σ	1,00	—	—	—	—	0,363069

Deci, mărimea medie a întreruperii în timpul de lucru al unui muncitor din colectivitatea generală ($\overline{X}$) va fi cuprinsă în intervalul :

$$\bar{x} \pm t \cdot \mu,$$

adică

$$30{,}41 \pm 1{,}18,$$

ceea ce înseamnă că mărimea medie a întreruperii în timpul de lucru zilnic al unui muncitor din întreaga colectivitate de 1 200 lucrători nu va fi mai mică de 29,23 minute și nici mai mare de 31,59 minute și aceasta, cu o siguranță probabilistică de 95%.

2° *Cazul cînd sînt studiate două caracteristici.* Formulele care dau precizia selecției tipice nerepetate ne sînt cunoscute. Le vom adapta acum pentru cazul cînd unitățile colectivității generale au fiecare cîte două caracteristici a și b. Fie

$$u_{1j}, u_{2j}, \ldots, u_{n_j, j}$$
$$v_{1j}, v_{2j}, \ldots, v_{n_j, j}$$

valorile caracteristicilor u și v luate de unitățile selectate din grupa j.

Media lor, $\bar{u}_j$ și $\bar{v}_j$, este

$$\bar{u}_j = \frac{\sum_{k=1}^{n_j} u_{kj}}{n_j}, \qquad \bar{v}_j = \frac{\sum_{k=1}^{n_j} v_{kj}}{n_j}. \qquad \text{(XX, 7.7)}$$

Referitor la aceeași grupă, notăm

$$\tilde{s}^2_{a_j a_j} = \frac{1}{n_j - 1}\sum_{i=1}^{n_j}(u_{ij} - \bar{u}_j)^2,$$

$$\tilde{s}^2_{a_j b_j} = \frac{1}{n_j - 1}\sum_{i=1}^{n_j}(u_{ij} - \bar{u}_j)(v_{ij} - \bar{v}_j), \qquad \text{(XX, 7.8)}$$

$$\tilde{s}^2_{b_j b_j} = \frac{1}{n_j - 1}\sum_{i=1}^{n_j}(v_{ij} - \bar{v}_j)^2.$$

În aceste condiții, admițînd că avem k grupe, expresiile

$$\bar{u} = \frac{1}{N}\sum_{j=1}^{k} N_j \bar{u}_j, \qquad \bar{v} = \frac{1}{N}\sum_{j=1}^{k} N_j \bar{v}_j \qquad \text{(XX, 7.9)}$$

ne dau niște estimații ale mediilor caracteristicilor a și b pentru întreaga colectivitate.

Dispersiile valorilor $\bar{u}$ și $\bar{v}$ sînt date respectiv de formulele:

$$D^2(\bar{u}) = \sum_{j=1}^{k} \frac{N_j^2}{N^2}(1 - f_j)\frac{\bar{s}^2_{a_j a_j}}{n_j}, \qquad \text{(XX, 7.10)}$$

$$D^2(\bar{v}) = \sum_{j=1}^{k} \frac{N_j^2}{N^2}(1 - f_j)\frac{\bar{s}^2_{b_j b_j}}{n_j}, \qquad \text{(XX, 7.11)}$$

$$c(\bar{u}, \bar{v}) = \sum_{j=1}^{k} \frac{N_j^2}{N^2}(1 - f_j)\frac{\bar{s}^2_{a_j b_j}}{n_j}, \qquad \text{(XX, 7.12)}$$

unde prin $c(\bar{u}, \bar{v})$ am notat covarianta variabilelor u și v.

Obținem, astfel, ca și în cazul selecției simple, precizia cu care determinăm mediile caracteristicilor a și b pentru colectivitatea întreagă.

Considerăm exemplul care urmează:

Într-o regiune care are 1 000 000 de locuitori ne propunem, în vederea unei acțiuni de școlarizare, ca pe baza unei anchete statistice să determinăm procentul de mecanici care au urmat o școală de specialitate față de numărul total al mecanicilor. Ancheta se bazează pe o selecție de 6% din populația totală. În mediul rural, în comunele cu mai puțin de 2 000 de locuitori, numărul mecanicilor este foarte redus. De aceea așezăm locuitorii din aceste comune într-o primă grupă din care efectuăm o selecție numai din 1% din populație. O a doua grupă o formăm din numărul locuitorilor din comunele avînd o populație cuprinsă între 2 000 și 5 000 de locuitori; o a treia grupă cu o populație cuprinsă între 5 001 și 15 000 locuitori și o a patra grupă cu peste 15 000.

Procedînd astfel întocmim următorul plan al selecției (tabelul 91).

Tabelul 91

Grupe *)	N_j	Procentul selecției	Populația selectată
a	1	2	3
Grupa I (0—2,000)	600 000	1%	6 000
Grupa II (2 001—5 000)	120 000	5%	6 000
Grupa III (5 001—15 000)	80 000	10%	8 000
Grupa IV (peste 15 000)	200 000	20%	40 000
	1 000 000		60 000

*) În parantezele fiecărei grupe am așezat numărul locuitorilor din comunele care alcătuiesc grupa, ca un indicativ general, ținînd seama că unele din aceste comune au putut fi trecute în grupe superioare.

Coloana a doua din tabel arată populația din fiecare grupă. În coloana a treia este trecut procentul selecției, care variază de la 1% pînă la 20%. După cum arată coloana ultimă, populația selectată se ridică la 60 000 persoane, deci la aceeași populație la care am fi ajuns dacă nu am fi procedat prin stratificare și am fi aplicat un procent general de selecție de 6%.

Să admitem că pentru exemplul considerat, rezultatele selecției pe grupe sînt:

Grupa	Mecanici (b)	Mecanici cu școală de specializare (a)
I	30	7
II	54	12
III	96	23
IV	523	136
	703	178

Caracteristica a, în cazul de față, reprezintă mecanicii cu o școală de specializare, iar caracteristica b — mecanici. Numerele 178 și 703 variază de la selecție la selecție. Dacă am fi luat altă selecție de 6% din colectivitatea generală, în loc de numerele de mai sus am fi găsit altele. Intervalul probabil de variație a numerelor de mai sus se calculează cu ajutorul dispersiei.

Să presupunem că, efectuînd calculele, am găsit că intervalul de variație pentru $u = 178$ este 178 ± 16, iar pentru $v = 703$ este 703 ± 30. Ne interesează, în baza acestor date, variația raportului

$$\frac{u}{v} = \frac{178}{703} = 25{,}32\%.$$

Aplicăm formulele (XX, 2.31), unde:

$$\bar{u} = 178, \quad \bar{v} = 703, \quad \delta = 16, \quad \varepsilon = 30.$$

Efectuînd calculele găsim

$$\frac{\delta}{\bar{v}} + \frac{\bar{u}\varepsilon}{\bar{v}^2} = 3{,}35\%.$$

Raportul $\frac{\bar{u}}{\bar{v}}$, cu o probabilitate mai mare decît 0,78, variază între 22% și 29%.

Rezultatul corect pe baza selecției efectuate este că, procentual, numărul mecanicilor cu o școală de specializare față de numărul total al mecanicilor, pe întreaga regiune, este cuprins între 22% și 29%.

Scanned with OKEN Scanner

8. Sondajul tipic proporțional sau reprezentativ

Particularitatea caracteristică a acestui fel de sondaj constă în faptul că din fiecare grupă tipică în care a fost despărțită colectivitatea generală se extrag, individual, atîtea unități astfel încît, raportul dintre numărul lor și volumul grupei din care s-au extras să fie egal cu raportul dintre volumul general al selecției și cel al colectivității generale.

Ne vom opri în continuare numai asupra sondajului tipic reprezentativ nerepetat.

Prin definiție, deci, se numește *sondaj tipic proporțional sau reprezentativ*, un sondaj grupat, simplu, pentru care sînt îndeplinite egalitățile

$$f_j = f, \qquad (j = 1, 2, \ldots, k)$$

de unde

$$f = \frac{n_1}{N_1} = \frac{n_2}{N_2} = \ldots = \frac{n_k}{N_k} = \frac{\sum_{j=1}^{k} n_j}{\sum_{j=1}^{k} N_j} = \frac{n}{N},$$

din care deducem

$$n_j = \frac{n}{N} N_j, \qquad (j = 1, 2, \ldots, k).$$

Prin urmare, într-un sondaj tipic reprezentativ după cum s-a mai spus, numărul unităților extrase din fiecare grupă este proporțional cu numărul unităților grupei.

Dispersia funcției de estimație y, în cazul sondajului reprezentativ, este

$$D^2(y) = \frac{1 - f}{n} \sum_{j=1}^{k} \frac{N_j}{N} s_j^2 \qquad \text{(XX, 8.1)}$$

ceea ce se poate ușor constata, introducînd valoarea de mai sus a lui n_j în expresia dispersiei din relațiile (XX, 7.4) și (XX, 7.5).

Să considerăm din nou exemplul colectivității generale de 1 200 muncitori în cadrul căreia, prin metoda sondajului tipic simplu, s-a luat în studiu mărimea medie a întreruperii în timpul zilnic de lucru al unui muncitor. De data aceasta, însă, vom aplica metoda sondajului tipic reprezentativ, notînd volumul general al sondajului ($n = 200$ sau 16,66% respectiv, 17%. Extrăgînd, deci, din fiecare grupă N_j aproximativ cîte 17% din numărul unităților respective, obținem următoarele sondaje de grupă (n_j) (tabelul 92).

Tabelul 92

Meseria (k)	Numărul muncitorilor extrași (n_j)	Meseria (k)	Numărul muncitorilor extrași (n_j)
Strungari	83 (41 %)	Ajustori	25 (13 %)
Frezori	34 (17 %)	Rectificatori	58 (29 %)

Abaterea mediei pătratică a mediei de selecție fiind egală cu rădăcina pătrată din expresia (XX, 8.1), adică

$$\mu = \sqrt{\frac{1-f}{n}\sum_{j=1}^{k}\frac{N_j}{N}s_j^2}, \qquad \text{(XX, 8.2)}$$

pentru determinarea ei avem nevoie de date suplimentare. Vom construi un tabel ajutător de calcul (tabelul 93).

Tabelul 93

Meseria $k_j (j = 1, 2, 3, 4)$	$\frac{N_j}{N}$	s_j^2	$\frac{N_j}{N} \cdot s_j^2$
Strungari	0,41	64	26,24
Frezori	0,17	100	17,00
Ajutori	0,13	49	6,37
Rectificatori	0,29	81	23,49
Σ	1,00	—	73,10

Datele din acest tabel le introducem în formula (XX, 8.2) și obținem :

$$\mu = \sqrt{\frac{1-\frac{200}{1\,200}}{200}\cdot 73{,}10} = \sqrt{0{,}304577} = 0{,}552 \text{ minute.}$$

După cum se vede, de data aceasta am obținut o eroare medie de selecție mai mică decît atunci cînd am aplicat metoda sondajului tipic simplu (0,552 față de 0,60).

Cu o probabilitate de 0,95 vom avea următoarea inegalitate

$$\bar{x} - (1{,}96 \cdot 0{,}552) \leqslant \bar{X} \leqslant \bar{x} + (1{,}96 \cdot 0{,}552).$$

Este evident că, întrucît acest rezultat s-a obținut cu același prag de probabilitate și cu același volum general al selecției ca și în primul caz, în cazul sondajului tipic reprezentativ lucrăm, în general cu mai multă precizie.

Trebuie totuși să facem o observație și anume : din punct de vedere teoretic nu se poate trage o concluzie generală în sensul că sondajul tipic simplu este mai puțin eficient decît sondajul tipic reprezentativ. Aceasta înseamnă că din punct de vedere teoretic se pot construi sondaje tipice simple care să fie mai eficiente decît sondajele tipice proporționale.

Se poate demonstra că pentru aceasta este suficient să aibă loc inegalitatea

$$n_j \leqslant N_j - \frac{N-n}{n}, \qquad (j = 1, 2, \ldots, k)$$

și toate valorile dispersiilor s_j să fie foarte mici.

Din punct de vedere practic, remarcăm că inegalitatea de mai sus nu este nici o dată satisfăcută deoarece ar trebui ca fiecare N_j să fie mai mic decît ordinul sutelor ($N_j \leqslant 100$) ceea ce nu se întîmplă în realitate, dacă ținem seama că metoda selecției nu se aplică decît atunci cînd numărul unităților colectivității generale este foarte mare.

O altă observație deosebit de importantă care trebuie făcută este aceea că *sondajul tipic proporțional este întotdeauna mai eficient decît sondajul simplu, întrucît dă naștere la o dispersie mai mică.*

Dacă considerăm următoarele două sondaje: primul un sondaj simplu de n unități efectuat din colectivitatea A, iar al doilea, un sondaj tipic proporțional, de aceeași mărime, efectuat din aceeași colectivitate A împărțită în k grupe, vom avea într-adevăr inegalitatea dispersiilor

$$\frac{1}{n}(1-f)s^2 \geqslant \frac{1}{n}(1-f)\sum_{j=1}^{n}\frac{N_j}{N}\cdot s_j^2, \qquad \text{(XX, 8.3)}$$

care exprimă tocmai faptul că sondajul tipic proporțional este mai eficient decît sondajul simplu.

Inegalitatea (XX, 8.3) se poate deduce astfel:

Avem pe de o parte

$$\sum_{j=1}^{k}\sum_{i=1}^{N_j}(a_{ij}-a)^2 = \sum_{j=1}^{k}\sum_{i=1}^{N_j}[(a_{ij}-a_j)+(a_j-a)]^2 =$$

$$= \sum_{j=1}^{k}\sum_{i=1}^{N_j}(a_{ij}-a_j)^2 + 2\sum_{j=1}^{k}\sum_{i=1}^{N_j}(a_{ij}-a_j)(a_j-a) + \sum_{j=1}^{k}N_j(a_j-a)^2 =$$

$$= \sum_{j=1}^{k}\sum_{i=1}^{N_j}(a_{ij}-a_j)^2 + 2\sum_{j=1}^{k}(a_j-a)\left(\sum_{i=1}^{N_j}a_{ij} - N_j a_j\right) + \sum_{j=1}^{k}N_j(a_j-a)^2 =$$

$$= \sum_{j=1}^{k}\sum_{i=1}^{N_j}(a_{ij}-a_j)^2 + \sum_{j=1}^{k}N_j(a_j-a)^2,$$

iar pe de altă parte,

$$s^2 = \frac{1}{N-1}\sum_{j=1}^{k}\sum_{i=1}^{N_j}(a_{ij}-a)^2$$

și

$$s_j^2 = \frac{1}{N_j-1}\sum_{i=1}^{N_j}(a_{ij}-a_j)^2.$$

Atunci

$$(N-1)s^2 = \sum_{j=1}^{k}(N_j-1)s_j^2 + \sum_{j=1}^{k}N_j(a_j-a)^2$$

sau fiindcă numerele N_j sînt foarte mari, putem să scriem egalitatea aproximativă

$$Ns^2 \simeq \sum_{j=1}^{k} N_j s_j^2 + \sum_{j=1}^{k} N_j (a_j - a)^2$$

deci

$$s^2 \geqslant \sum_{j=1}^{k} \frac{N_j}{N} s_j^2 .$$

Înmulțind ambii membri ai acestei inegalități cu numărul pozitiv $\frac{1}{n}(1-f)$, deducem tocmai inegalitatea (XX, 8.3).

9. Sondajul tipic optim

În teoria selecției se demonstrează că se poate determina un astfel de volum de selecție pe grupe $n_j, (j = 1, 2, \ldots, k)$ pentru care eficiența sondajului să fie maximă.

Aceasta revine la a face minimă expresia (XX, 7.4) condiționată de egalitatea

$$n_1 + n_2 + \ldots + n_k = n.$$

Se ajunge astfel la concluzia că trebuie să avem

$$n_j = \frac{n N_j s_j}{\sum_{j=1}^{k} N_j s_j}, \qquad (j = 1, 2, \ldots, k) \qquad \text{(XX, 9.1)}$$

Aceasta este deci, expresia care determină valoarea numerelor n_j, pentru care eficiența sondajului este maximă. Un asemenea sondaj, al cărui număr de unități n_j este dat de formula (XX, 9.1), poartă numele de sondaj stratificat optim. După cum se observă, pentru sondajul stratificat optim numărul unităților sondajului dintr-o grupă oarecare este direct proporțional cu numărul unităților din acea grupă și cu abaterea pătratică medie a grupei respective.

Pentru a face minimă funcția de selecție

$$D^2(y) = \sum_{j=1}^{k} \frac{N_j^2}{N^2}(1 - f_j)\frac{s_j^2}{n_j}$$

condiționată de egalitatea

$$\sum_{j=1}^{k} n_j = n,$$

se poate folosi metoda multiplicatorilor lui Lagrange.

Scanned with OKEN Scanner

Considerînd funcția

$$F(n_1, n_2, \ldots, n_k) = \sum_{j=1}^{k} \frac{N_j^2}{N^2}(1 - f_j)\frac{s_j^2}{n_j} + \lambda\left(\sum_{j=1}^{k} n_j - n\right) =$$

$$= \sum_{j=1}^{k} \frac{N_j^2 s_j^2}{N^2 n_j} - \sum_{j=1}^{k} \frac{N_j s_j^2}{N^2} + \lambda\left(\sum_{j=1}^{k} n_j - n\right)$$

și anulînd derivatele ei parțiale de ordinul întîi în raport cu n_j, obținem

$$\frac{\partial F(n_1, n_2, \ldots, n)_k}{\partial n_j} = -\frac{N_j^2 s_j^2}{N^2 n_j^2} + \lambda = 0, \quad (j = 1, 2, \ldots, k).$$

ceea ce arată că λ este pozitiv. De aici rezultă

$$\frac{N_j s_j}{N n_j} = \sqrt{\lambda}$$

luînd pentru radical semnul plus, deoarece numerele N, N_j, s_j și n_j sînt toate pozitive.

Egalitatea precedentă fiind echivalentă cu

$$\frac{\sum_{j=1}^{k} N_j s_j}{nN} = \sqrt{\lambda},$$

rezultă că

$$\frac{N_j s_j}{n_j} = \frac{\sum_{j=1}^{k} N_j s_j}{n},$$

de unde deducem expresia (XX, 9.1).

Pentru a cunoaște eficiența sondajului tipic optim este necesar să calculăm dispersia funcției de estimație (XX, 7.2) corespunzătoare acestui sondaj. Ea se obține înlocuind valoarea lui n_j din (XX, 9.1) în expresia $D^2(y)$ dată de (XX, 7.4). Astfel, găsim

$$D^2(y) = \sum_{j=1}^{k} \frac{N_j^2}{N^2}\left(1 - \frac{n_j}{N_j}\right)\frac{s_j^2}{n_j} = \sum_{j=1}^{k} \frac{N_j^2 s_j^2}{N^2}\left(\frac{1}{n_j} - \frac{1}{N_j}\right) =$$

$$= \sum_{j=1}^{k} \frac{N_j s_j^2}{N^2}\left(\frac{\sum_{j=1}^{k} N_j s_j}{n s_j} - 1\right) = \frac{1}{N^2}\left[\frac{\left(\sum_{j=1}^{k} N_j s_j\right)^2}{n} - \sum_{j=1}^{k} N_j s_j^2\right] \quad \text{(XX, 9.2)}$$

Să considerăm din nou exemplul folosit în cazul sondajului tipic simplu și proporțional, de data aceasta însă, efectuînd un sondaj optim, — alegem, în acest scop, numerele n_j potrivit formulei (XX, 9.1).

Vom calcula elementele ce ne sînt necesare pentru a stabili și în acest caz intervalul în care se va găsi media $\overline{X}$ în colectivitatea generală.

Tabelul 94

Meserii	$N_j s_j$	$\frac{N_j s_j}{\sum_{j=1}^{n} N_j s_j}$	$n_j = \frac{n N_j s_j}{\sum_{j=1}^{n} N_j s_j}$
Strungari	4 000	0,39	78
Frezori	2 000	0,19	38
Ajustori	1 050	0,11	22
Rectificatori	3 150	0,31	62
Total:	10 200	1,00	200

În cazul sondajului optim, abaterea medie pătratică a mediei de selecție va fi dată de formula:

$$\sqrt{D^2(\overline{x})} = \mu = \sqrt{\frac{1}{N^2}\left[\frac{\left(\sum_{j=1}^{k} N_j s_j\right)^2}{n} - \sum_{j=1}^{k} N_j s_j^2\right]}. \qquad (XX, 9.3)$$

Să determinăm valorile de sub radical:

Tabelul 95

Meserii	$(N_j s_j)^2$	s_j^2	$N_j s_j^2$
Strungari	16 000 000	64	32 000
Frezori	4 000 000	100	20 000
Ajustori	1 102 500	49	7 350
Rectificatori	9 922 500	81	28 350
Total:	31 025 000		87 700

Vom avea deci:

$$\mu = \sqrt{\frac{1}{1200^2}\left[\frac{31025000}{200} - 87\,700\right]} = \sqrt{\frac{1}{1200^2}\left[155125 - 87700\right]} =$$

$$= \sqrt{0{,}0447} \simeq 0{,}21 \text{ minute}$$

și

$$\overline{x} = \frac{1}{n}\sum_{j=1}^{k} n_j \overline{x}_j = \frac{(78 \cdot 30) + (38 \cdot 40) + (32 \cdot 20) + (62 \cdot 30)}{200} =$$

$$= \frac{6360}{200} = 31{,}8 \text{ minute}$$

Această medie de selecție este — după cum știm — o variabilă aleatoare, și de aceea, la un alt sondaj tipic optim, ea poate lua o altă valoare (de pildă, 30, 76, sau 30, 81 etc), variind însă între anumite limite, care, cu o probabilitate de 0,95 nu vor fi depășite și anume:

$$\bar{x} \pm 1{,}96, \mu \text{ adică } 31{,}8 \pm 1{,}96 \cdot 0{,}21.$$

Aceasta înseamnă că mărimea medie a întreruperii în timpul de lucru zilnic al unui muncitor din colectivitatea generală nu va fi mai mică de 31,4 minute și nici mai mare de 32,2 minute; această afirmație evident, o facem, cu probabilitatea de 0,95, aplicînd legea normală.

După cum se vede din exemplul considerat, *sondajul tipic optim este mai eficient decît sondajul tipic proporțional* și se poate demonstra că această afirmație are caracterul unei *teoreme generale*, întrucît dispersia mediei de selecție în cazul sondajului tipic optim este întotdeauna mai mică decît dispersia în cazul sondajului tipic proporțional.

Demonstrația acestei teoreme rezultă nemijlocit din faptul că diferența dintre dispersiile corespunzătoare ale celor două sondaje, date de relațiile (XX, 7.4), respectiv, (XX, 9.2), este întotdeauna un număr pozitiv.

Într-adevăr, avem relația

$$\frac{1-f}{n}\sum_{j=1}^{k}\frac{N_j}{N}s_j^2 - \frac{1}{N^2}\left[\frac{\left(\sum_{j=1}^{k}N_j s_j\right)^2}{n} - \sum_{j=1}^{k}N_j s_j^2\right] = \frac{1}{n}\left[\sum_{j=1}^{k}\frac{N_j s_j^2}{N} - \left(\sum_{j=1}^{k}\frac{N_j s_j}{N}\right)^2\right] =$$

$$= \frac{1}{n}\left[\sum_{j=1}^{k}\frac{N_j}{N}\left(s_j - \sum_{j=1}^{k}\frac{N_j s_j}{N}\right)^2\right] > 0$$

ceea ce demonstrează teorema enunțată.

Dispersia sondajului tipic proporțional nu poate fi egală cu dispersia sondajului tipic optim decît în ipoteza cînd toate numerele s_j, $(j = 1, 2, \ldots, k)$ sînt egale între ele, adică numai dacă dispersiile de grupă sînt aceleași.

Economicitatea sondajului tipic optim. Este evident că din punct de vedere științific sondajul tipic optim este cel mai exact. Problema se pune însă și din punctul de vedere al economicității lui urmărindu-se în realitate să se obțină maximum de informație statistică cu un anumit cost. Cunoaștem astfel cheltuielile c_j, $(j = 1,2,\ldots,k)$, pe o singură unitate din fiecare grupă A_j, $(j = 1,2,\ldots,k)$.

După cum se vede, presupunem că aceste cheltuieli variază de la grupă la grupă, ceea, ce, în general, corespunde realității. De pildă, s-ar putea, dacă cele două grupe reprezintă mediul rural și mediul urban, ca unele anchete statistice selective să coste pentru fiecare unitate în mod diferit, de la un mediu la altul. Dispunem pentru aplicarea metodei selecției de o sumă totală C, pe care n-o putem depăși. Se pune atunci problema determinării unităților n_j corespunzătoare sondajului din fiecare grupă, astfel ca să obținem cea mai mare eficiență.

Aceasta înseamnă să stabilim minimul expresiei (XX, 7.4), adică

$$D^2(\bar{y}) = \sum_{j=1}^{k}\frac{N_j^2}{N^2}(1-f_j)\frac{s_j^2}{n_j}$$

condiționat de egalitatea

$$\sum_{j=1}^{k} n_j c_j = C$$

care exprimă faptul că suma totală a cheltuielilor este C.

Un calcul simplu ne arată că trebuie să avem

$$n_j = \frac{N_j s_j}{\sqrt{c_j}} \cdot \frac{C}{\sum_{j=1}^{k} N_j s_j \sqrt{c_j}} \qquad (j = 1, 2, \ldots, k). \qquad \text{(XX, 9.4)}$$

Am determinat astfel numărul unităților din fiecare grupă, în așa fel încît să avem cea mai mare eficiență a sondajului.

Expresia (XX, 9.4) se poate deduce astfel: Aplicînd metoda multiplicatorilor lui Lagrange, avem de considerat funcția

$$F(n_1, n_2, \ldots, n_k) = \sum_{j=1}^{k} \frac{N_j^2}{N^2}(1 - f_j)\frac{s_j^2}{n_j} + \lambda\left(\sum_{j=1}^{k} n_j c_j - C\right) =$$

$$= \sum_{j=1}^{k} \frac{N_j^2 s_j^2}{N^2 n_j} - \sum_{j=1}^{k} \frac{N_j s_j^2}{N^2} + \lambda\left(\sum_{j=1}^{k} n_j c_j - C\right)$$

Anulînd derivatele ei parțiale de ordinul întîi în raport cu n_j, obținem

$$\frac{\partial F(n_1, n_2, \ldots, n_k)}{\partial n_j} = -\frac{N_j^2 s_j^2}{N^2 n_j^2} + \lambda c_j = 0, \qquad (j = 1, 2, \ldots, k),$$

unde se vede de asemenea că $\lambda > 0$.

De aci rezultă

$$n_j = \frac{N_j s_j}{N\sqrt{\lambda} \cdot \sqrt{c_j}},$$

unde $\sqrt{\lambda}$ s-a luat cu semnul plus, deoarece n_j este, evident, un număr pozitiv.

Pe de altă parte, avînd în vedere relația

$$\sum_{j=1}^{k} n_j c_j = C,$$

deducem

$$C = \frac{\sum_{j=1}^{k} N_j s_j \sqrt{c_j}}{N\sqrt{\lambda}}$$

adică

$$\frac{1}{N\sqrt{\lambda}} = \frac{C}{\sum_{j=1}^{k} N_j s_j \sqrt{c_j}}.$$

Prin urmare, expresia lui n_j de mai sus devine tocmai expresia (XX, 9.4)

Din relația (XX, 9.4), prin însumare în raport cu indicatorul j de la 1 la k, deducem imediat mărimea sondajului, care este

$$n = C \sum_{j=1}^{k} \frac{N_j s_j}{\sqrt{c_j} \sum_{j=1}^{k} N_j s_j \sqrt{c_j}}. \qquad \text{(XX, 9.5)}$$

În ceea ce privește mărimea dispersiei care exprimă cea mai mare eficiență a sondajului, tipic, în condițiile problemei date, ea capătă următoarea expresie:

$$D^2(y) = \sum_{j=1}^{k} \frac{N_j^2}{N^2}(1-f_j)\frac{s_j^2}{n_j} = \sum_{j=1}^{k} \frac{N_j^2 s_j^2}{N^2}\left(\frac{1}{n_j} - \frac{1}{N_j}\right) =$$

$$= \sum_{j=1}^{k} \frac{N_j s_j^2}{N^2}\left(\frac{\sqrt{c_j}\sum_{j=1}^{k} N_j s_j \sqrt{c_j}}{C s_j} - 1\right) = \frac{1}{N^2}\left(\frac{\left[\sum_{j=1}^{k} N_j s \sqrt{c_j}\right]^2}{C} - \sum_{j=1}^{k} N_j s_j^2\right). \qquad \text{(XX, 9.6)}$$

În particular, dacă

$$c_1 = c_2 = \ldots = c_k = c$$

și deci,

$$C = c \sum_{j=1}^{k} n_j = cn,$$

regăsim, după cum era și firesc, expresiile (XX, 9.1) și (XX, 9.2) ale lui n_j și $D^2(y)$ pentru cazul sondajului tipic optim.

Într-adevăr, relațiile (XX, 9.4) și (XX, 9.6) se pot atunci scrie

$$n_j = \frac{N_j s_j}{c} \cdot \frac{nc}{\sum_{j=1}^{k} N_j s_j} = \frac{nN_j s_j}{\sum_{j=1}^{k} N_j s_j}$$

și

$$D^2(y) = \frac{1}{N^2}\left[\frac{c\left(\sum_{j=1}^{k} N_j s_j\right)^2}{nc} - \sum_{j=1}^{k} N_j s_j^2\right] = \frac{1}{N^2}\left[\frac{\left(\sum_{j=1}^{k} N_j s_j\right)^2}{n} - \sum_{j=1}^{k} N_j s_j^2\right].$$

Problema economicității sondajului se pune și din alt punct de vedere. De pildă, presupunînd că se cunoaște eficiența sondajului general, σ^2, se cere să se determine numărul n_j al sondajului din fiecare grupă $A_j, (j = 1, 2, \ldots, k)$, astfel ca cheltuielile totale C să fie minime. Cu alte cuvinte trebuie să găsim minimul expresiei

$$C = \sum_{j=1}^{k} n_j c_j \qquad \text{(XX, 9.7)}$$

condiționat de egalitatea

$$\sigma^2 = \sum_{j=1}^{k} \frac{N_j^2}{N^2}(1-f_j)\frac{s_j^2}{n_j}.$$

Obținem, efectuînd calculele,

$$n_j = \frac{N_j s_j}{\sqrt{c_j}} \cdot \frac{\sum_{j=1}^{k} N_j s_j \sqrt{c_j}}{N\sigma^2 + \sum_{j=1}^{k} N_j s_j^2}. \qquad (XX, 9.8)$$

În ambele probleme de interes economic pe care le-am semnalat, se vede, din examinarea relațiilor (XX, 9.4) și (XX, 9.7) că numărul unităților din fiecare grupă este direct proporțional cu numărul unităților N_j din acea grupă și cu abaterea medie pătratică s_j a grupei respective și invers proporțional cu rădăcina pătrată a cheltuielilor c_j pe o singură unitate din grupa corespunzătoare.

Pentru a calcula pe n_j, vom aplica de asemenea metoda multiplicatorilor lui Lagrange, problema fiind de a calcula acum minimul funcției:

$$F(n_1, n_2, \ldots, n_k) = \sum_{j=1}^{k} n_j c_j + \lambda \left[\sum_{j=1}^{k} \frac{N_j^2}{N^2} (1 - f_j) \frac{s_j^2}{n_j} - \sigma^2 \right] =$$

$$= \sum_{j=1}^{k} n_j c_j + \lambda \left[\sum_{j=1}^{k} \frac{N_j^2 s_j^2}{N^2 n_j} - \sum_{j=1}^{k} \frac{N_j s_j^2}{N^2} - \sigma^2 \right].$$

Luînd derivatele parțiale de ordinul întîi în raport cu n_j ale funcției $F(n_1, n_2, \ldots, n_k)$ și anulîndu-le găsim

$$\frac{\partial F(n_1, n_2, \ldots, n_k)}{\partial n_j} = c_j - \lambda \frac{N_j^2 s_j^2}{N^2 n_j^2} = 0, \qquad (j = 1, 2, \ldots, k)$$

din care deducem

$$n_j = \frac{N_j s_j \sqrt{\lambda}}{N \sqrt{c_j}},$$

semnul $\sqrt{\lambda}$ fiind evident pozitiv.

Atunci, expresia lui σ^2, care arată eficiența sondajului, devine

$$\sigma^2 = \sum_{j=1}^{k} \frac{N_j^2 s_j^2}{N^2 n_j} - \sum_{j=1}^{k} \frac{N_j s_j}{N^2} = \frac{1}{N\sqrt{\lambda}} \sum_{j=1}^{k} N_j s_j \sqrt{c_j} - \frac{1}{N^2} \sum_{j=1}^{k} N_j s_j^2.$$

Din această egalitate putem afla ușor pe $\sqrt{\lambda}$ care este egal cu

$$\sqrt{\lambda} = \frac{N \sum_{j=1}^{k} N_j s_j \sqrt{c_j}}{N^2\sigma^2 + \sum_{j=1}^{k} N_j s_j^2}.$$

Scanned with OKEN Scanner

În sfîrșit, introducînd valoarea lui $\sqrt{\lambda}$ în expresia lui n_j de mai sus, obținem numărul unităților sondajului din fiecare grupă pentru care cheltuielile totale sînt minime, adică tocmai relația (XX, 9.8)

$$n_j = \frac{N_j s_j}{\sqrt{c_j}} \cdot \frac{\sum_{j=1}^{k} N_j s_j \sqrt{c_j}}{N\sigma^2 + \sum_{j=1}^{k} N_j s_j^2}, \qquad (j = 1, 2, \ldots, k).$$

Avînd valorile lui n_j, calculul cheltuielilor minime rezultă imediat și anume:

$$C = \sum_{j=1}^{k} n_j c_j = \sum_{j=1}^{k} N_j s_j \sqrt{c_j} \cdot \frac{\sum_{j=1}^{k} N_j s_j \sqrt{c_j}}{N\sigma^2 + \sum_{j=1}^{k} N_j s_j^2} = \frac{\left(\sum_{j=1}^{k} N_j s_j \sqrt{c_j}\right)^2}{N\sigma^2 + \sum_{j=1}^{k} N_j s_j^2}.$$

Din cele spuse mai sus cu privire la problemele de economicitate a sondajului tipic optim, cititorul s-a putut convinge că nivelul cheltuielilor totale sau pe unitate de observare trebuie avut întotdeauna în vedere în procesul determinării volumului acestui fel de sondaj. Calculul volumului sondajelor de grupă și al sondajului general capătă, în acest caz, alte forme. Pentru a-l efectua corect, statisticianul trebuie să mai țină seama — după cum am văzut — de principiul științific de constituire a grupelor tipice, de caracterul unităților de observare, de eroarea limită de selecție ca și de siguranța probabilistică dinainte stabilită. Toți acești factori sînt considerați în strînsă legătură cu volumul cheltuielilor pe care le necesită executarea sondajului în ansamblu sau în volumul cheltuielilor ce revine pe unitate de observare.

Socotim că aplicarea corectă a metodologiei cuprinsă în expresiile (XX, 9.4), (XX, 9.5), (XX, 9.6) și (XX, 9.8) asigură economicitatea maximă a sondajului tipic.

*

Problemele tratate în cazul sondajului tipic optim se pun și în ipoteza cînd unitățile din colectivitatea generală A, au mai multe caracteristici.

În cazul sondajului cu mai multe caracteristici, problema eficienții sondajului devine însă mult mai complicată.

Să examinăm, pentru a pune în evidență dificultatea problemei, cazul unui sondaj cu două caracteristici.

Mai întîi trebuie definită eficiența. Astfel, dacă avem două caracteristici u și v, se poate, pe baza unei analize a fenomenului considerat, să dăm întîietate caracteristicii u și să ne intereseze să facem minimă dispersia $D^2(\bar{u})$. Problema în acest caz se reduce la cazul unei singure caracteristici. Repetînd același raționament se vede că trebuie să luăm

$$n_j = \frac{n_j N_j s_{a_j} a_j}{\sum_{j=1}^{k} N_j s_{a_j} a_j}.$$

Prin urmare, cînd dăm întîietate unei caracteristici problema este simplă.

Se poate însă ca ambele caracteristici să ne intereseze în egală măsură. Cu alte cuvinte, aplicînd metoda selecției ne propunem să obținem precizia maximă pentru ambele caracteristici.

În acest caz, va trebui să alegem numerele n_j în așa fel încît aria elipsei de concentrație a variabilelor $\bar{u}$ și $\bar{v}$ să fie minimă.

Plecăm de la matricea covariantă a variabilelor $\bar{u}$ și $\bar{v}$:

$$\begin{Vmatrix} D^2(\bar{u}) & c(\bar{u}, \bar{v}) \\ c(\bar{u}, \bar{v}) & D^2(\bar{v}), \end{Vmatrix}$$

unde prin $c(\bar{u}, \bar{v})$ am notat covarianța variabilelor $\bar{u}$ și $\bar{v}$.

Dacă notăm

$$A = \frac{D^2(\bar{v})}{D^2(\bar{u})\,D^2(\bar{v}) - c^2(\bar{u}, \bar{v})},$$

$$B = -\frac{c(\bar{u}, \bar{v})}{D^2(\bar{u})D^2(\bar{v}) - c^2(\bar{u}, \bar{v})},$$

$$C = \frac{D^2(\bar{u})}{D^2(\bar{u})D^2(\bar{v}) - c^2(\bar{u}, \bar{v})},$$

elipsoidul de concentrație are ecuația :

$$Ax^2 - 2Bxy + cy^2 = 4.$$

Aria acestei elipse este

$$\frac{4\pi}{AC - B^2} = 4\pi[D^2(\bar{u})\,D^2(\bar{v}) - c^2(\bar{u}, \bar{v})].$$

Pentru ca aria să fie minimă, trebuie deci ca expresia $D^2(\bar{u})\,D^2(\bar{v}) - C^2(\bar{u}, \bar{v})$ să fie minimă.

Ținînd seama de formulele (XX, 7.10), (XX,7,11), (XX, 7.12) și aplicînd metoda multiplicatorilor lui Lagrange, trebuie să calculăm minimul funcției

$$\sum_{j=1}^{k} \frac{N_j^2}{N^2}(1 - f_j)\,\frac{s^2_{a_j a_j}}{n_j} \sum_{j=1}^{k} \frac{N_j^2}{N^2}(1 - f_j)\,s^2_{b_j b_j} - \left(\sum_{j=1}^{k} \frac{N_j}{N^2}(1 - f_j)\,\frac{s^2_{a_j b_j}}{n_j}\right)^2 + \lambda\left(\sum_{j=1}^{k} n_j - n\right).$$

Derivatele parțiale ale acestei funcții în raport cu n_j, $(j = 1, 2, \ldots, k)$, ne conduc la un sistem care nu poate fi rezolvat decît în anumite cazuri particulare.

10. Cîteva observații generale asupra sondajului tipic

Am mai arătat că sondajul tipic (simplu proporțional și optim), folosind combinat principiile grupării științifice cu principiile selecției, asigură cele mai mici erori de reprezentativitate în comparație cu celelalte feluri de selecții și prin aceasta, calitatea mai bună a indicatorilor sintetici estimați ai colectivității generale. Variația caracteristicilor de selecție obținute pe baza sondajului, conține, în general, influența a două componente :

a) variația produsă în cadrul diferitelor grupe sub influența factorilor obiectivi întîmplători, specifici acelor grupe tipice, variația *internă* de grupă și

b) variația *dintre grupe*, reflectînd media abaterilor mediilor de grupă de la media sondajului general.

Aplicînd sondajul tipic noi ne eliberăm de variația *dintre grupe tipice*, astfel încît, la extinderea rezultatelor selecției pentru a estima indicatorii colectivității generale, această variație nu-și exercită influența asupra erorii limită de selecție. Tocmai aceasta este și explicația faptului că în formulele funcțiilor de estimație a dispersiei $D^2(y)$ (vezi (XX, 7.4) (XX, 8.1) (XX, 9.2), în cazul sondajului tipic nu se mai folosește dispersia caracteristicii din colectivitatea generală, ci *media dispersiilor de grupă* care reflectă, în esență, variația suferită de caracteristica studiată în cadrul fiecărei grupe, dacă, în toate grupele, ar fi acționat aceiași factori obiectivi întîmplători.

Cu totul altfel stau lucrurile cînd se aplică sondajul simplu. Asupra erorii întîmplătoare de reprezentativitate în acest fel de sondaj am văzut că exercită influență variația caracteristicii la toate unitățile sondajului luat în ansamblu. Este clar că dacă variația internă din cadrul fiecărei grupe tipice este mai mică decît variația pe întregul sondaj, aplicarea selecției tipice duce la o micșorare bine determinată a erorii limită și în consecință, la sporirea preciziei indicatorilor de selecție.

De aici decurge o concluzie de importanță practică deosebită. Pentru statisticianul însărcinat cu organizarea unei selecții tipice se pune în primul rînd problema de a forma grupe care, prin conținutul lor, să fie cu adevărat tipice sub aspectul caracteristicilor studiate. Rezolvînd această problemă în condiții optime el va asigura omogenitatea deplină a fiecărei grupe în parte, omogenitatea la rîndul ei va fi caracterizată prin valori medii tipice substanțial deosebite, de la o grupă la alta.

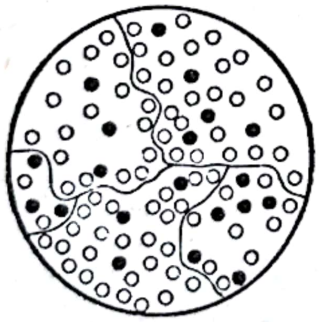


Fig. 46

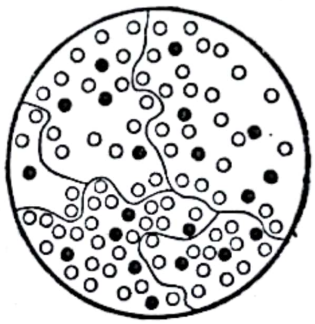


Fig. 47

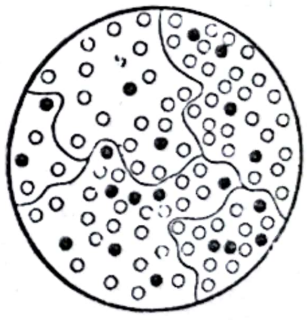


Fig. 48

În paragrafele precedente noi ne-am oprit numai asupra selecțiilor tipice, nerepetate, construite după schema probabilistică hipergeometrică care, după cum se știe, în comparație cu schema probabilistică a bilei revenite, este caracterizată printr-o dispersie a variabilei mai mică de $\frac{N-n}{N-1} \cong 1 - \frac{n}{N}$ ori.

Dacă practic se impune executarea unei selecții tipice repetate, formulele folosite pentru stabilirea abaterii medii pătratice, a mediei de selecție, a erorii limită de selecție și a volumului sondajului vor avea

aceeași formă cu singura deosebire că va dispare factorul de micșorare a dispersiei $(1 - f_i)$.

Sondajul tipic simplu poate fi reprezentat grafic ca în figura 46.

Un exemplu de asemenea sondaj îl poate oferi distribuția bugetelor de familie muncitorești pe regiuni.

Sondajul tipic proporțional poate fi reprezentat grafic ca în figura 47.

Un asemenea sondaj este folosit la noi în studiul bugetelor de familie în vederea cunoașterii nivelului de trai al muncitorilor, țăranilor colectiviști și individuali, precum și al funcționarilor. El a căpătat caracterul unei lucrări statistice permanente, devenind un instrument deosebit de prețios în conducerea planificată a economiei naționale.

Sondajul tipic optim poate fi reprezentat ca în figura 48.

Scanned with OKEN Scanner

Scanned with OKEN Scanner

Capitolul XXI

Sondaje în două sau mai multe faze şi alte sondaje

1. Sondajul în două faze, generalităţi

Denumirea de sondaj în două faze se atribuie de obicei, acelor cercetări selective în care unităţile de selecţie de acelaşi tip constituie obiectul observării în două faze.

Sondajul în două faze se obţine alegînd în prima fază $h < k$ grupe la întîmplare dintr-o colectivitate A împărţită în k straturi.

$$A_1, A_2, \ldots, A_k$$

şi apoi efectuînd, în a doua fază, din cele h grupe ieşite, extracţii în acelaşi mod ca la sondajul tipic simplu (colectivitatea generală şi componenţa grupelor admitem că sînt aceleaşi ca la sondajul tipic simplu). Sondajul tipic simplu este un caz particular al sondajului în două faze, pentru $h = k$.

Aceleaşi probleme care se pun pentru sondajul tipic simplu se pun şi pentru sondajul în două faze, care se impune de preferinţă atunci cînd trebuie să se efectueze o statistică într-un timp mai scurt sau cînd suma totală a cheltuielilor ar fi prea mare dacă sondajul s-ar face într-o singură fază.

De exemplu, să presupunem că într-o ţară sînt 10 000 de comune rurale. Pe baza unui sondaj se cere evaluarea numărului de familii cu mai mult de 5 copii din comunele rurale. Considerînd că fiecare comună rurală reprezintă o grupă, putem să estimăm numărul de familii pe care-l căutăm, făcînd, de pildă, la întîmplare cîte un sondaj de 10% din fiecare comună rurală. Procedînd astfel, aplicăm metoda sondajului stratificat

simplu. Această metodă necesită, însă, un timp de lucru mai îndelungat, deoarece conduce la un număr de unități mai mare.

De aceea se poate proceda la un sondaj în două faze, alegînd la întîmplare, în prima fază, 100 de comune rurale, de pildă, din totalul de 10 000 de comune rurale.

Apoi, în a doua fază, din fiecare dintre aceste 100 comune rurale alegem la întîmplare, de pildă, un procent de 10%. Se vede, deci, utilitatea sondajului în două faze din punct de vedere economic.

În practică s-au folosit sondaje în două faze și în alt scop, caracteristică fiind următoarea situație: în prima fază sînt studiate toate unitățile colectivității de selecție în ceea ce privește anumite caracteristici bine determinate, în faza a doua, însă, pot fi cercetate numai o parte din unitățile care au fost extrase în prima fază și aceasta cu scopul de a fi studiate unele caracteristici suplimentare. Așa, de pildă, dacă în prima fază a unei selecții au fost studiate 5 000 de familii, înregistrîndu-se cu acest prilej o singură caracteristică, să zicem nivelul venitului fiecărei familii, în a doua fază se poate alege la întîmplare 10% din cele 5 000 de familii cercetate în prima fază, înregistrîndu-se însă acum și alte caracteristici suplimentare cum ar fi cele referitoare la sursa venitului, destinația lui pe cheltuieli etc.

Rezolvarea matematică a problemei sondajului în două faze este aceeași într-un caz și în altul.

2. Sondajul în două faze cu o singură caracteristică

Întocmai ca la sondajul cu o singură fază și o singură caracteristică să notăm prin

$$x_{ij}, \qquad (i = 1, 2, \ldots, n_j)$$

valorile ieșite pentru unitățile secundare din grupa A_j, $(j = 1, 2, \ldots, k)$ și prin Y_j expresia

$$Y_j = \frac{kN_j}{N} \frac{\sum_{i=1}^{n_j} x_{ij}}{n_j}. \qquad (XXI, 2.1)$$

Fie

$$Y = \frac{Y_{j_1} + Y_{j_2} + \ldots + Y_{j_k}}{k}, \qquad (XXI, 2.2)$$

$j_1, j_2, \ldots, j_k$ arătînd indicii grupelor care ne-au ieșit la întîmplare în executarea primei faze a sondajului.

Întrebuinţînd aceleaşi notaţiuni ca pînă acum, putem enunţa următoarea :

TEOREMĂ. *Într-un sondaj cu două faze avem :*

$$M(Y) = a, \qquad \text{(XXI, 2.3)}$$

$$D^2(Y) = \frac{k(k-h)}{hN^2}\,\frac{\sum_{j=1}^{k}(\alpha_j-\alpha)^2}{k-1} + \frac{k}{hN^2}\sum_{j=1}^{k} N_j^2(1-f_j)\frac{s_j^2}{n_j}, \qquad \text{(XXI, 2.4)}$$

unde

$$a = \frac{1}{N}\sum_{j=1}^{k}\sum_{i=1}^{N_j} a_{ij},$$

$$\alpha_j = \sum_{i=1}^{N_j} a_{ij},$$

$$\alpha = \frac{1}{k}\sum_{j=1}^{k}\alpha_j$$

iar f_j şi s_j^2 avînd aceleaşi semnificaţii ca în teoremele precedente.

Demonstraţia acestei teoreme se obţine uşor dacă ţinem seamă de rezultatele obţinute pînă acum la sondajul cu o singură fază.

Probabilitatea ca din h extracţii efectuate la întîmplare din cele k unităţi primare să obţinem grupările :

$$A_{j_1}, A_{j_2} \ldots, A_{j_h}$$

în ordinea menţionată mai sus, este $\frac{h!}{k^l}$.

Notăm prin

$$Y_{j_1}, Y_{j_2} \ldots, Y_{j_h}$$

nişte variabile aleatoare independente, care corespund efectiv alegerii straturilor de mai sus.

Fie de asemenea

$$F_j(x) = P(Y_j \leqslant x), \qquad (j = 1, 2, \ldots, k)$$

funcţia de repartiţie a variabilei Y_j.

Probabilitatea ca să avem

$$x_{j_1} < Y_{j_1} < x_{j_1} + \mathrm{d}x_{j_1},\ x_{j_2} < Y_{j_2} < x_{j_2} + \mathrm{d}x_{j_2}, \ldots,\ x_{j_h} < Y_{j_h} < x_{j_h} + \mathrm{d}x_{j_h}$$

este

$$\frac{h!}{k^l}\,\mathrm{d}F_{j_1}(x_{j_1})\,\mathrm{d}F_{j_2}(x_{j_2}) \ldots \mathrm{d}F_{j_h}(x_{j_h}).$$

De aici rezultă

$$M(Y) = M\left(\frac{(Y_{j_1} + Y_{j_2} + \ldots + Y_{j_h})}{h}\right) =$$

$$= \frac{h!}{k^l}\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\cdots\int_{-\infty}^{+\infty}\frac{x_{j_1} + x_{j_2} + \ldots + x_{j_h}}{h}\,\mathrm{d}F_{j_1}(x_{j_1}) \ldots \mathrm{d}F_{j_h}(x_h) =$$

$$= \frac{h!}{k^l}\sum_{j_1, j_2, \ldots, j_h}\frac{m_{j_1} + \ldots + m_{j_h}}{h} = \frac{1}{k}\sum_{j=1}^{k} m_j = m.$$

În egalitățile de mai sus am notat

$$M(Y_j) = m_j, \qquad (j = 1, 2, \ldots, k)$$

și

$$m = \frac{1}{k} \sum_{j=1}^{k} m_j.$$

De asemenea am ținut seama de faptul că în suma

$$\sum_{j_1, \ldots, j_h} \frac{m_{j_1} + \ldots + m_{j_h}}{h},$$

coeficientul unui m_{j_s} determinat este $\frac{(k-1)!}{(h-1)!} \cdot \frac{1}{h}$.

În mod analog obținem

$$M(Y^2) = \frac{h!}{k!} \sum_{j_1, \ldots, j_h} \int_{-\infty}^{+\infty} \int_{\cdots}^{+\infty} \int_{-\infty}^{+\infty} \frac{(x_{j_1} + x_{j_2} + \ldots + x_{j_h})^2}{h^2} \, dF_{j_1}(x_{j_1}) \ldots dF_{j_h}(x_{j_h}) =$$

$$= \frac{h!}{k!} \sum_{j_1, \ldots, j_h} \int_{-\infty}^{+\infty} \int_{\cdots}^{+\infty} \int_{-\infty}^{+\infty} \frac{\sum_{v=1}^{h} x_{j_v}^2 + 2 \sum_{v \neq s} x_{j_v} x_{j_s}}{h^2} \, dF_{j_1}(x_{j_1}) \ldots dF_{j_h}(x_{j_h}) =$$

$$= \frac{h!}{k!} \sum_{j_1, \ldots, j_h} \frac{\sum_{v=1}^{h} M_{j_v} + 2 \sum_{v \neq s} m_{j_v} m_{j_s}}{h^2} = \frac{h!}{k!} \sum_{j_1, \ldots, j_h} \frac{\sum_{v=1}^{h} (M_{j_v} - m_{j_v}^2) + \left(\sum_{v=1}^{h} m_{j_v} \right)^2}{h^2},$$

unde am notat

$$M(Y_j^2) = M_j, \qquad (j = 1, 2, \ldots, k).$$

Din această relație și din relația precedentă, rezultă

$$D^2(Y) = M(Y^2) - [M(Y)]^2 =$$

$$= \frac{h!}{k!} \sum_{j_1, \ldots, j_h} \frac{\sum_{v=1}^{h} (M_{j_v} - m_{j_v}^2)}{h^2} + \frac{h!}{k!} \sum_{j_1, \ldots, j_h} \left[\left(\frac{\sum_{v=1}^{h} m_{j_v}}{h} \right)^2 - \left(\frac{\sum_{j=1}^{k} m_j}{k} \right)^2 \right].$$

În prima sumă coeficientul unui termen $M_{j_s} - m_{j_s}^2$, pentru un indice j_s determinat, este

$$\frac{h!}{k!} \frac{(k-1)!}{(h-1)!} \frac{1}{h^2} = \frac{1}{hk}.$$

Prin urmare, avem

$$\frac{h!}{k!} \sum_{j_1, \ldots, j_s} \frac{\sum_{v=1}^{h} (M_{j_v} - m_{j_v}^2)}{h^2} = \frac{1}{hk} \sum_{j=1}^{k} (M_j - m_j^2) = \frac{k}{hN^2} \sum_{j=1}^{k} N_j^2 (1 - f_j) \frac{s_j^2}{n_j},$$

deoarece de la sondajul simplu știm că

$$M_j - m_j^2 = D^2(Y_j) = \frac{k^2}{N^2} N_j^2 (1 - f_j) \frac{s_j^2}{n_j}.$$

Suma a doua se poate scrie :

$$\frac{h!}{k!}\sum_{j_1,\ldots,j_h}\left[\left(\frac{\sum_{j=1}^{h} m_{j_v}}{h}\right)^2-\left(\frac{\sum_{j=1}^{h} m_j}{k}\right)^2\right]=\frac{h!}{k!}\sum_{j_1,\ldots,j_h}\left[\left(\frac{\sum_{v=1}^{h} m_{j_v}}{h}\right)^2-m^2\right]=$$

$$=\frac{1}{hk(k-1)}\left[(k-1)\sum_{j=1}^{k} m_j^2+2(h-1)\sum_{i\neq j} m_i m_j-hk(k-1)m^2\right]=\frac{k-h}{hk}\,\frac{\sum_{j=1}^{k}(m_j-m)^2}{k-1}=$$

$$=\frac{k-h}{kh}\,\frac{\sum_{j=1}^{k}\left(\frac{k}{N}N_j a_j-\frac{1}{k}\sum_{j=1}^{k}\frac{k}{N}N_j a_j\right)^2}{k-1}=\frac{k(k-h)}{hN^2}\cdot\frac{\sum_{j=1}^{k}(\alpha_j-\alpha)^2}{k-1}.$$

Am obținut astfel expresia lui $M(Y)$ și $D^2(Y)$.

Teorema de mai sus ne dă suportul teoretic pentru aplicarea în practică a sondajului în două faze, deoarece valoarea medie a lui Y este tocmai media aritmetică a caracteristicilor pentru întreaga colectivitate A. Pentru ca rezultatele obținute să fie exacte, este necesar, în virtutea inegalității lui Cebîșev, ca dispersia lui Y dată de (XXI, 2.4) să fie foarte mică. Vom examina cazul cînd acest lucru se întîmplă.

După cum se constată, dispersia variabilei aleatoare Y este suma a doi termeni :

$$D_1=\frac{k(k-h)}{hN^2}\,\frac{\sum_{j=1}^{k}(\alpha_j-\alpha)^2}{k-1}.$$

și

$$D_2=\frac{k}{hN^2}\sum_{j=1}^{k} N_j^2\,(1-f_j)\,\frac{s_j^2}{n_j}.$$

Primul termen D_1 reprezintă dispersia corespunzătoare primei faze a sondajului și se referă la suma caracteristicilor unităților secundare,

$$\alpha_j=\sum_{i=1}^{N_j}\alpha_{ij}.$$

Deci, putem considera acest termen ca o dispersie între grupe. Acest termen se anulează cînd $h=k$, adică atunci cînd sondajul încetează de a mai fi cu două faze și se reduce la un sondaj grupat simplu. De asemenea, se mai anulează, și atunci cînd

$$\alpha_1=\alpha_2=\ldots=\alpha_k,$$

cu alte cuvinte, atunci cînd suma caracteristicilor unităților secundare este aceeași pentru fiecare grupă.

Al doilea termen D_2 reprezintă dispersia din interiorul grupelor care corespund fazei a doua a sondajului. El se anulează cînd avem $n_j=N_j$, $(j=1,2,\ldots,k)$, ceea ce înseamnă atunci cînd scoatem toate unitățile secundare din fiecare grupă.

Rezultă de aici că dacă numerele n_j sînt suficient de mari, termenul $D_2\to 0$, în timp ce termenul D_1 rămîne constant (deoarece nu depinde de n_j).

Prin urmare, variabila aleatoare

$$Y = \frac{Y_{j_1} + Y_{j_2} + \ldots + Y_{j_h}}{h}$$

nu este o funcție de estimație pentru media caracteristicilor colectivității totale a, oricît de mare ar fi mărimea

$$n = \sum_{j=1}^{k} n_j$$

a sondajului.

Pentru ca din punct de vedere practic să ajungem la rezultate acceptabile printr-un sondaj în două faze este necesar ca dispersia $D^2(Y)$ să fie foarte apropiată de zero pentru numerele n_j, $(j = 1, 2, \ldots, k)$ suficient de mari. Aceasta însă, după cele arătate mai sus, nu se poate obține decît numai dacă D_1 este, la rîndul său, foarte apropiat de zero. Însă ca D_1 să fie egal cu zero, trebuie ca

$$h = k$$

sau

$$\alpha_j = \alpha, \qquad (j = 1, 2, \ldots, k).$$

Primul caz nu interesează, deoarece avantajul sondajului în două faze constă tocmai în faptul de a alege la întîmplare un număr mai mic de unități primare decît cele existente.

Rămîne deci de rezolvat numai cazul cînd

$$\alpha_j = \alpha, \qquad (j = 1, 2, \ldots, k),$$

de unde rezultă că spre deosebire de sondajul grupat simplu, sondajul în două faze nu este aplicabil decît după o prealabilă prelucrare a grupelor.

Pentru a pune în evidență erorile ce se pot ivi în aplicarea fără un control prealabil al sondajului în două faze, dăm următorul exemplu, redus la o formă schematică, simplă:

Presupunem astfel că avem:

$$k = 4, \qquad h = 2,$$

adică un sondaj în patru grupe din care am ales la întîmplare cîte două.

Admitem că pentru fiecare grupă

$N_1 = 10\,000$	$\alpha_1 = 100\,000$
$N_2 = 12\,000$	$\alpha_2 = 25\,000$
$N_3 = 10\,000$	$\alpha_3 = 200\,000$
$N_4 = 18\,000$	$\alpha_4 = 75\,000$
$N = 50\,000$	$\alpha = 400\,000$

Valorile α_j diferă sensibil între ele.

Media caracteristicilor colectivității totale este

$$a = \frac{\alpha}{N} = 8.$$

Să mai admitem, după ce am ales cele două unități primare, că lucrăm în cele mai exacte condiții, adică din fiecare grupă scoatem toate unitățile secundare. În aceste condiții, aplicînd formulele stabilite rezultă pentru media a următoarele estimații :

Faza a doua a sondajului s-a făcut din grupele	Estimația mediei a
A_1, A_2	5
A_1, A_3	12
A_1, A_4	7
A_2, A_3	9
A_2, A_4	4
A_3, A_4	11

Valoarea exactă a mediei a fiind 8, se vede de aici că aplicarea sondajului în două faze în cele mai optime condiții ne conduce la rezultate inacceptabile din punct de vedere practic.

Să presupunem că, păstrînd aceleași unități primare și secundare, avem :

$N_1 = 10\,000$	$\alpha_1 = 100\,000$
$N_2 = 12\,000$	$\alpha_2 = 90\,000$
$N_3 = 10\,000$	$\alpha_3 = 110\,000$
$N_4 = 18\,000$	$\alpha_4 = 100\,000$
$N = 50\,000$	$\alpha = 400\,000$

În acest caz, valorile α_i nu diferă prea mult între ele și deci ne așteptăm la estimații mult mai apropiate de realitate.

Media caracteristicilor colectivității totale este aceeași :

$$a = \frac{\alpha}{N} = 8.$$

Lucrînd în aceleași condiții, adică alegînd la întîmplare două grupe din cele patru și extrăgînd din fiecare grupă unitățile secundare, obținem pentru media a estimațiile de mai jos :

Faza a doua a sondajului s-a făcut de grupele	Estimația mediei a
A_1, A_2	7,6
A_1, A_3	8,8
A_1, A_4	8,0
A_2, A_3	8,0
A_2, A_4	7,6
A_3, A_4	8,8

Scanned with OKEN Scanner

Aceste rezultate sînt mult mai apropiate de valoarea exactă. Este evident că dacă am lua toate valorile

$$\alpha_i = 100\,000, \qquad (\alpha_i = 1, 2, 3, 4)$$

estimația mediei a ar fi chiar egală cu 8.

Această observație are valoare practică, deoarece ne arată că pentru aplicarea unui sondaj în două faze este necesar să ne alegem grupele astfel, încît suma caracteristicilor unităților din fiecare grupă să fie aproximativ aceeași.

Abaterea medie patratică pentru sondajul în două faze cu o singură caracteristică va fi dată, evident, de rădăcina pătrată din expresia (XXI, 2.4). Se poate conchide că eroarea sondajului în două faze rezultă din suma erorilor celor două faze ale acestuia și că este mai mare decît oricare dintre acestea. Prin urmare, în comparație cu celelalte feluri de selecție, sondajul în două faze este mai puțin precis dat fiind că în faza a doua sporește eroarea totală. Sondajul în două faze are însă, după cum s-a mai spus, o serie de avantaje din punct de vedere organizatoric și al economicității, mai ales atunci cînd obiectul cercetării are o mare extindere teritorială.

Un alt avantaj al sondajului în faze constă în aceea că mediile se pot obține pe fiecare grupă inclusă în sondaj, la fel ca și în cazul sondajelor tipice. De relevat este că selecția în faze poate reduce atît erorile de înregistrare cît și erorile de reprezentativitate grosolane, asigurînd, de aceea, obținerea unor concluzii de multe ori mai juste decît în cazul aplicării altor procedee de selecție.

Numărul exagerat al fazelor sondajului trebuie însă evitat, știut fiind că fiecare fază nouă mărește eroarea totală.

Într-adevăr, eroarea medie de selecție a unui sondaj în mai multe faze este dată de formula:

$$\mu = \sqrt{\frac{k(k-h)}{hN^2}\,\frac{\sum_{j=1}^{k}(\alpha_j-\alpha)^2}{k-1} + \frac{k}{hN^2}\sum_{j=1}^{k} N_j^2\,(1-f_j)\,\frac{s_j^2}{n_j}}\,.$$

3. Sondajul în două faze cu mai multe caracteristici

Se examinează problema mai generală a sondajului în două faze pentru cazul cînd unitățile fiecărui strat A_j, $(j = 1,2,\ldots,k)$, prezintă fiecare cîte două caracteristici.

Unitățile secundare sînt respectiv vectorii aleatori

$$h_{ij}(a_{ij}, b_{ij}), \qquad (i = 1, 2, \ldots, N_j), \qquad (j = 1, 2, \ldots, k),$$

iar unitățile corespunzătoare fazei a doua a sondajului sînt, respectiv, vectorii aleatori:

$$x_{ij}(u_{ij}, v_{ij}), \qquad (i = 1, 2, \ldots, n_j), \qquad (j = 1, 2, \ldots, k).$$

Fie atunci

$$Y_j = \frac{kN_j}{N} \cdot \frac{\sum_{i=1}^{n_j} x_{ij}}{n_j},$$

vectorul aleator care are componentele u_j și v_j și

$$Y = \frac{Y_{j_1} + Y_{j_2} + \ldots + Y_{j_h}}{h}$$

vectorul aleator cu componentele

$$U = \frac{u_{j_1} + u_{j_2} + \ldots + u_{j_h}}{h}$$

și

$$V = \frac{v_{j_1} + v_{j_2} + \ldots + v_{j_h}}{h}.$$

Toate notațiile introduse la studiul sondajului tipic simplu cu două caracteristici rămîn valabile și pentru sondajul în două faze avînd două caracteristici.

TEOREMĂ. *Într-un sondaj cu două faze, avînd două caracteristici, există relațiile*

$$M(U) = a, \quad M(V) = b,$$

$$D^2(U) = \frac{k(k-h)}{hN^2} \cdot \frac{\sum_{j=1}^{k} (\alpha_j - \alpha)^2}{k-1} + \frac{k}{hN^2} \sum_{j=1}^{k} N_j^2 (1-f_j) \frac{s^2_{a_j a_j}}{n_j},$$

$$D^2(V) = \frac{k(k-h)}{hN^2} \cdot \frac{\sum_{j=1}^{k} (\beta_j - \beta)^2}{k-1} + \frac{k}{hN^2} \sum_{j=1}^{k} N_j^2 (1-f_j) \frac{s^2_{b_j b_j}}{n_j}, \qquad \text{(XXI, 3.1)}$$

$$C(UV) = \frac{k(k-h)}{hN^2} \cdot \frac{\sum_{j=1}^{k} (\alpha_j - \alpha)(\beta_j - \beta)}{k-1} + \frac{k}{hN^2} \sum_{j=1}^{k} N_j^2 (1-f_j) \frac{s^2_{a_j b_j}}{n_j},$$

unde am notat prin $C(UV) = M(U-a)(V-b)$ *covarianța variabilelor* U *și* V, *iar*

$$\alpha_j = \sum_{i=1}^{N_j} a_{ij}, \quad \beta_j = \sum_{i=1}^{N_j} b_{ij}, \quad \alpha = \frac{1}{k} \sum_{j=1}^{k} \alpha_j \quad \text{și} \quad \beta = \frac{1}{k} \sum_{j=1}^{k} \beta_j.$$

Primele patru formule (XXI, 3.1) se stabilesc întocmai ca la sondajul în două faze cu o singură caracteristică.

Probabilitatea ca în prima fază a sondajului să obținem grupele:

$$A_{j_1}, \quad A_{j_2}, \ldots, \quad A_{j_h}$$

este $\frac{h!}{k!}$.

Fie

$$F_j(x, y) = P(U_j \leqslant x, \; V_j \leqslant y), \qquad (j = 1, 2, \ldots, k)$$

funcția de repartiție a vectorului $Y_j(U_j, V_j)$.

Atunci, probabilitatea de a avea

$$x_{j_1} < U_{j_1} < x_{j_1} + \mathrm{d}x_{j_1}, \; x_{j_2} < U_{j_2} < x_{j_2} + \mathrm{d}x_{j_2}, \ldots, x_{j_h} < U_{j_h} < x_{j_h} + \mathrm{d}x_{j_h}$$

și

$$y_{j_1} < V_{j_1} < y_{j_1} + \mathrm{d}y_{j_1}, \; y_{j_2} < V_{j_2} < y_{j_2} + \mathrm{d}y_{j_2}, \ldots, y_{j_h} < V_{j_h} < y_{j_h} + \mathrm{d}y_{j_h}$$

este egală cu

$$\frac{h!}{k!}\,dF_{j_1}(x_{j_1}, y_{j_1})\,dF_{j_2}(x_{j_2}, y_{j_2})\ldots dF_{j_h}(x_{j_h}, y_{j_h}),$$

de unde rezultă

$$M(UV)=\frac{h!}{k!}\sum_{j_1 j_2 \ldots j_h}\iint\ldots\int\frac{x_{j_1}+x_{j_2}+\ldots+x_{j_h}}{h}\times$$

$$\times\frac{y_{j_1}+y_{j_2}+\ldots+y_{j_h}}{h}\,dF_{j_1}(x_{j_1}, y_{j_1}).\ dF_{j_2}(x_{j_2}, y_{j_2})\ldots dF_{j_h}(x_{j_h}, y_{j_h}), \qquad \text{(XXI, 3.2)}$$

fiecare dintre integrale extinzîndu-se în tot planul.

Să mai introducem notația

$$m_j^{(1)} = \int x_j\,dF_j(x_j, y_j),$$

$$m_j^{(2)} = \int y_j\,dF_j(x_j, y_j)$$

și

$$M_{ij} = \int x_j\,y_j\,dF_j(x_j, y_j),$$

unde fiecare dintre aceste integrale se extinde, în mod analog, în tot planul. Cu aceste notații, relația (XXI, 3.2) devine

$$M(UV)=\frac{h!}{k!}\sum_{j_1, j_2, \ldots, j_h}\frac{\sum_{r=1}^{h}M_{j_r j_r}+\sum_{r\neq s}m_{j_r}^{(1)}m_{j_s}^{(2)}}{h^2}=$$

$$=\frac{h!}{k!}\sum_{j_1, j_2, \ldots, j_h}\frac{\sum_{r=1}^{h}(M_{j_r j_r}-m_{j_r}^{(1)}m_{j_r}^{(2)})+\sum_{r=1}^{h}m_{j_r}\sum_{s=1}^{h}m_{j_s}}{h^2}. \qquad \text{(XXI, 3.3)}$$

Pe de altă parte, avem

$$M(U)\cdot M(V)=\frac{\sum_{j=1}^{k}m_j^{(1)}}{k}\cdot\frac{\sum_{j=1}^{k}m_j^{(2)}}{k}. \qquad \text{(XXI, 3.4)}$$

Din relațiile (XXI, 3.3) și (XXI, 3.4) deducem

$$C(UV)=M(UV)-M(U)\,M(V)=\frac{h!}{k!}\sum_{j_1, j_2, \ldots, j_h}\frac{\sum_{r=1}^{h}(M_{j_r j_r}-m_{j_r}^{(1)}m_{j_r}^{(2)})}{h^2}+$$

$$+\frac{h!}{k!}\sum_{j_1, \ldots, j_h}\left[\frac{\sum_{r=1}^{h}m_{j_r}}{h}\cdot\frac{\sum_{r=1}^{h}m_{j_s}}{h}-\frac{\sum_{j=1}^{k}m_j^{(1)}}{k}\cdot\frac{\sum_{j=1}^{k}m_j^{(2)}}{k}\right]= \qquad \text{(XXI, 3.5)}$$

$$=\frac{\sum_{j=1}^{k}(M_{jj}-m_j^{(1)}m_j^{(2)})}{hk}+\frac{k-h}{hk}\cdot\frac{\sum_{j=1}^{k}(m_j^{(1)}-m^{(1)})(m_j^{(2)}-m^{(2)})}{k-1},$$

avînd în vedere formulele stabilite la studiul sondajului simplu cu două caracteristici şi însemnînd

$$m^{(1)} = \frac{1}{k}\sum_{j=1}^{k} m_j^{(1)}, \quad m^{(2)} = \frac{1}{k}\sum_{j=1}^{k} m_j^{(2)}.$$

Observăm că, în cazul de faţă, putem scrie

$$M_{jj} - m_j^{(1)} m_j^{(2)} = \frac{k^2 N_j^2}{N^2}(1-f_j)\frac{s^2_{a_j b_j}}{n_j}$$

şi

$$(m_j^{(1)} - m^{(1)})(m_j^{(2)} - m^{(2)}) = \frac{k^2(\alpha_j - \alpha)(\beta_j - \beta)}{N^2};$$

rezultă, prin înlocuire în (XXI, 3.5), ultima formulă din relaţiile (XXI, 3.1) din teorema enunţată, adică

$$C(UV) = \frac{k(k-h)}{hN^2} \cdot \frac{\sum_{j=1}^{k}(\alpha_j - \alpha)(\beta_j - \beta)}{k-1} + \frac{k}{hN^2} \cdot \sum_{j=1}^{k} N_j^2 (1-f_j)\frac{s^2_{a_j b_j}}{n_j}.$$

Toate problemele care se pun în legătură cu sondajul cu o singură fază se pun şi pentru sondajul în două faze.

De pildă, ne putem propune estimarea raportului $\frac{a}{b}$ dintre mediile caracteristicilor colectivităţii generale precum şi problema determinării unui sondaj optim, ţinînd seama de eficienţa sondajului sau de cheltuielile totale.

Aceste probleme nu le vom trata; ele pot fi soluţionate prin metode analoge cu cele folosite la studiul sondajului cu o singură fază.

Scanned with OKEN Scanner

Capitolul XXII

Sondajul cu volum redus. Concluzii asupra metodei selecţiei

1. Consideraţii generale

Am examinat pînă acum cazurile cînd, în general, condiţiile concrete permit ca n (volumul subcolectivităţii de selecţie), să fie destul de mare, ceea ce asigură un grad mai mare de precizie a rezultatelor selecţiei. În anumite situaţii însă, nu este posibilă o selecţie cu volum mare şi totuşi metoda selecţiei se impune ca fiind singura posibilă din punct de vedere practic. Aşa, de pildă, la studierea statistică prin metoda selecţiei a calităţii producţiei, chiar în timpul procesului de producţie, pentru a preveni rebuturile, nu este practic posibil să se ia în studiu un volum mare de produse, pentru că controlul necesită rezultate imediate, operative, condiţie ce nu se poate îndeplini dacă sondajul are un volum mare, deoarece sondajul constă în alegerea la intervale egale de timp a produselor fabricate pentru a le măsura anumite caracteristici prevăzute în proiectul tehnic de fabricaţie. Se impune deci, în mod firesc, necesitatea de a fi efectuat un sondaj în care numărul unităţilor observate să nu depăşească valorile de 10—20. Un asemenea sondaj este denumit sondaj de volum mic.

Un număr mic de cazuri observate este caracteristic îndeosebi pentru normarea tehnică. Pentru determinarea lor după criterii ştiinţifice este necesar să se cunoască consumul specific de muncă pe o anumită operaţie şi aceasta, în medie, pe întreaga colectivitate de muncitori sau, tot în medie, pe o parte a ei, pe baza cronometrării timpului de lucru. Practic, consumul specific de timp de muncă pe operaţii se înregistrează numai la un număr mic de muncitori. Se face deci un sondaj cu volum redus.

Pentru estimarea rezultatelor sondajului cu volum redus nu ne putem însă folosi de valorile distribuţiei normale $\Phi(t)$ ale mediilor de selecţii, întrucît — după cum am mai arătat — numai dacă volumul sondajului este foarte mare, rezultatele date de cercetarea selectivă sînt pre-

cise, iar distribuția mediilor de selecție posibile se face după legea normală. În cazul unui număr mic de experiențe, caracterul distribuției mediei de selecție depinde încă prea puternic de caracterul distribuției mărimilor individuale. De aceea, dacă nu știm nimic referitor la distribuția acestor mărimi, nu putem preciza nici caracterul distribuției mediei lor. Însă, dacă cunoaștem caracterul distribuțiilor mărimilor considerate, putem face anumite aprecieri și asupra distribuției mediei lor. În particular, știm că (vezi Cap. VII) dacă distribuția fiecărei mărimi este normală, atunci și distribuția sumei lor și a mediei lor este de asemenea normală.

2. Verificarea ipotezelor

La sondajele cu volum redus există cîteva distribuții mai importante care se aplică de obicei. Să presupunem că avem de verificat exactitatea unei ipoteze statistice (H_0).

Această ipoteză poate fi de mai multe feluri, de exemplu :

Dacă avem un șir de frecvențe $x_1, x_2, \ldots, x_n$ reprezentînd o selecție de volum n, ne putem întreba dacă distribuția teoretică a colectivității generale este normală sau nu. Ipoteza că colectivitatea teoretică este normală poate fi considerată ca ipoteză H_0.

Putem să ne punem și o altă ipoteză. De pildă, putem să ne întrebăm dacă frecvențele de mai sus sînt realizări empirice ale distribuției teoretice

$$\frac{1}{\sqrt{2\pi}} e^{-\frac{(x-1)^2}{2}}.$$

Cu alte cuvinte, ne putem întreba, de data aceasta, nu numai dacă distribuția teoretică este normală, ci și dacă media este 1 și dispersia de asemenea 1.

Prin urmare, există numeroase ipoteze statistice, care însă toate se verifică după același procedeu și anume :

Ipoteza se încadrează într-un proces aleator, în care ea trebuie să se verifice cu o probabilitate foarte mare $p = 1 - \alpha$ unde α este un număr foarte mic. În acest caz din cauză că p este foarte aproape de 1, este firesc ca ipoteza să se verifice. Admitem astfel că ipoteza a fost adevărată. În cazul contrar vom respinge ipoteza, afirmînd că este neadevărată. Admiterea sau respingerea ipotezii o facem cu pragul de probabilitate α pe care l-am ales. Trebuie să facem observația esențială că prin procedee statistice nu putem afirma în mod cert că o ipoteză este adevărată sau neadevărată. O asemenea concluzie nu o putem face, decît numai cu o probabilitate foarte mare, ceea ce, totuși, din punct de vedere teoretic nu reprezintă certitudinea, însă poate să ne satisfacă din punct de vedere practic.

Admițînd sau respingînd o ipoteză statistică putem face erori de două feluri :

— putem respinge o ipoteză justă ; aceasta este o eroare de primul gen.

— putem admite o ipoteză falsă. Aceasta este o eroare de genul al doilea.

Scanned with OKEN Scanner

Alegînd un test, putem, prin mărirea volumului selecției, să facem ca probabilitatea comiterii unei erori de un anumit gen să fie arbitrar de mică. Nu putem însă să facem același lucru pentru probabilitățile erorilor de ambele genuri. Astfel dacă ne alegem un test și facem volumul selecției atît de mare, încît probabilitatea ca să putem obține o eroare de genul întîi, ar fi, de pildă, egală cu 0,05, în acest caz probabilitatea comiterii unei erori de genul al doilea se prezintă ca o consecință; ea nu mai stă la libera noastră dispoziție. Invers, putem fixa probabilitatea comiterii unei erori de genul al doilea și atunci rezultă ca o consecință probabilitatea comiterii unei erori de genul întîi. Nu se poate afirma care dintre aceste probabilități trebuie să fie mai mică sau mai mare, deoarece nu există o regulă generală în această privință.

De exemplu, dacă dorim să verificăm componența unui medicament, care peste un anumit grad de concentrare devine vătămător, comiterea unei erori de genul al doilea este mult mai primejdioasă decît comiterea unei erori de genul întîi. În schimb, la acceptarea unui articol de îmbrăcăminte, care nu corespunde întocmai condițiilor de fabricație, eroarea de primul gen este mai gravă decît eroarea de genul al doilea.

3. Distribuția χ^2

Criteriul lui χ^2 datorit lui K. A. Pearson, are largă aplicație în verificarea ipotezelor statistice. Vom arăta cum se aplică acest criteriu folosind cîteva exemple.

Să considerăm o succesiune de variabile independente $x_i, (i=1,2,\ldots,n)$, care pot lua valorile a_i $(1 \leqslant i \leqslant s)$, cu probabilitățile respective p_i $(1 \leqslant i \leqslant s)$

$$p_i \geqslant 0, \quad \sum_{i=1}^{s} p_i = 1.$$

Să notăm $\lambda_1, \lambda_2, \lambda_3, \ldots, \lambda_m$ respectiv numerele care arată de cîte ori s-au produs valorile $a_1, a_2, \ldots, a_m$ într-o experiență efectuată pentru primele n variabile x_i:

$$\lambda_i \geqslant 0, \sum_{i=1}^{s} \lambda_i = n.$$

Am văzut în capitolul VII că legea de probabilitate a variabilei

$$\sum_{i=1}^{s} \frac{(\lambda_i - np_i)^2}{np_i} = \chi^2$$

este

$$\frac{\chi^{s-2} e^{-\frac{\chi^2}{2}} d\chi}{2^{\frac{s-3}{2}} \Gamma\left(\frac{(s-1}{2}\right)}.$$

După cum am arătat, aceasta este o distribuție asimptotică de probabilitate. Deci, trebuie aplicată pentru valorile mari ale lui n.

Rezultă atunci că

$$p(\chi^2 \leqslant \chi_0^2) = \frac{1}{2^{\left(\frac{s-3}{2}\right)}\Gamma\left(\frac{s-1}{2}\right)} \int_0^{\chi_0^2} \chi^{s-2} e^{-\frac{\chi^2}{2}}\, d\chi.$$

Ne propunem să verificăm ipoteza H că probabilitățile $p_1, p_2, \ldots, p_s$ au valorile $\bar{p}_1, \bar{p}_2, \ldots \bar{p}_s$ respective.

Pentru aceasta calculăm valoarea lui χ^2 cu ajutorul relației

$$\sum_{i=1}^{s} \frac{(\lambda_i - n\bar{p}_i)^2}{n\bar{p}_i} = \bar{\chi}^2. \qquad \text{(XXII,3.1)}$$

Să alegem pe χ_0 astfel încît $(\chi^2 < \chi_0^2) = 1-\alpha$ unde $1-\alpha$ este un număr apropiat de 1. Deci cu o probabilitate foarte mare, trebuie să avem

$$\bar{\chi}^2 \leqslant \chi_0^2$$

dacă totuși

$$\chi^2 > \chi_0^2$$

ipoteza trebuie să fie respinsă.

EXEMPLUL 1. Presupunem că într-o regiune geografică oarecare situația privind numărul locuitorilor și numărul anual al deceselor este cea arătată în tabelul 96 :

Tabelul 96

Vîrsta	Numărul locuitorilor	Numărul deceselor într-un an
0	1	2
25	623 748	4 836
30	586 732	5 032
35	564 475	4 704
40	534 238	7 120
45	489 384	7 857
50	453 248	9 402

După o tabelă de mortalitate teoretică, probabilitățile teoretice sînt date de tabelul 97.

Tabelul 97

Vîrsta	Probabilitatea de deces q_χ
0	1
25	0,007 971
30	0,008 352
35	0,008 319
40	0,012 339
45	0,016 158
50	0,020 509

Scanned with OKEN Scanner

Pentru calculul lui χ^2 construim tabelul 98.

Tabelul 98

Vîrsta	Numărul anual al decedaților (l_i)	Numărul teoretic al decedaților (t_i)	$\frac{(l_i - t_i)^2}{t_i}$
0	1	2	3
25	4 836	4 972	3,720
30	5 032	4 900	3,555
35	4 704	4 695	0,011
40	7 120	7 153	0,152
45	7 857	7 907	0,316
50	9 402	9 295	1,231
			8,985

În coloana întîi a acestui tabel sînt trecute valorile empirice corespunzătoare fiecărei vîrste, adică numărul anual al decedaților constatat prin observări statistice. În coloana a doua sînt trecute numerele teoretice corespunzătoare valorilor din coloana întîia. De exemplu conform tabelului 97, probabilitatea de deces pentru vîrsta de 25 de ani este:

$$q_{25} = 0{,}007\,971.$$

Din tabelul 96 rezultă că numărul locuitorilor de 25 de ani este de 623 748. Prin urmare, numărul mediu anual teoretic al decedaților în vîrstă de 25 ani este

$$623\,748\, q_{25} = 623\,784 \cdot 0{,}007\,971 = 4\,972.$$

În același mod se calculează și celelalte elemente ale coloanei a doua din tabelul 98.

În ultima coloană din tabelul 98 este dată valoarea lui $\frac{(l_i - t_i)^2}{t_i}$. După cum se vede, prin adunarea tuturor numerelor din ultima coloană, obținem pentru χ^2 valoarea 8,985.

Căutînd în tabelele numerice ale lui χ^2 (vezi tabela-anexă II) găsim că, pentru șase grade de libertate, cîte avem în cazul nostru — deoarece există șase valori ale lui x — probabilitatea ca să avem o valoare a lui χ^2, mai mare decît 12,592, este 0,05. Intervalul $\chi^2 > 12{,}592$ este pentru noi intervalul critic, în care dacă este cuprinsă o valoare a lui χ^2, rezultă că valorile furnizate de experiență se îndepărtează în mod semnificativ de datele teoretice considerate. Această concluzie se impune pentru considerentul că este greu să admitem că în baza unei experiențe găsim tocmai o valoare care se prezintă în numai 5% din cazuri. Valoarea obținută pentru χ^2 este, în exemplul nostru, 8,985. Ea este cu mult mai mică decît 12,6 și deci nu poate fi vorba de o concluzie în genul celei de mai sus. Rezultă deci și pe această cale că mortalitatea efectivă din tabelul 96 nu diferă semnificativ de mortalitatea teoretică dată de tabelul 98.

EXEMPLUL 2. Să presupunem că se cunoaște numărul locuitorilor și numărul decedaților din două regiuni. Se poate deduce pe baza acestor

date că mortalitatea populațiilor din cele două regiuni diferă semnificativ. Pentru regiunea A avem datele din tabelul 99.

Tabelul 99

Vîrsta	Numărul locuitorilor	Numărul anual al decedaților
0	1	2
15—19	65 343	439
20—24	62 539	475
25—29	60 131	491
30—34	58 114	490
35—39	55 231	566
40—44	52 171	725
45—49	48 132	859
50—54	43 659	979
55—59	38 342	1 028

Mortalitatea pentru regiunea B rezultă din tabelul 100.

Tabelul 100

Vîrsta	Numărul locuitorilor	Numărul anual al decedaților
0	1	2
15—19	97 538	605
20—24	94 548	671
25—29	89 932	755
30—34	87 031	740
35—39	82 537	834
40—44	78 048	1 108
45—49	72 032	1 239
50—54	66 072	1 454
55—59	57 031	1 551

Pentru rezolvarea acestei probleme vom folosi criteriul lui χ^2. Vom arăta cum se calculează valoarea lui χ, pentru prima grupă de vîrstă 15—19 ani. Considerăm tabelul 101.

Tabelul 101

Regiunea	Numărul supraviețuitorilor după un an	Numărul anual al decedaților	Total
0	1	2	3
A	64 904	439	65 343
B	96 933	605	97 538
	161 837	1 044	162 881

În prima coloană sînt dați indicatorii privind numărul locuitorilor rămași în viață după un an. În coloana a doua, numărul anual al decedaților. Datele sînt scoase din tabelele 99 și 100, pentru grupa de vîrstă 15—19 ani. Coloana a treia conține totalul elementelor corespunzătoare din fiecare rînd, iar rîndul trei conține totalul elementelor corespunzătoare din fiecare coloană.

Ne propunem să verificăm ipoteza că mortalitatea populației este aceeași în ambele regiuni. În acest caz, populația totală (la începutul anului) cuprinde 162 881 locuitori dintre care 161 837 au rămas în viață după un an și 1 044 au murit.

Probabilitatea ca un locuitor să trăiască peste un an este

$$p = \frac{161\,837}{162\,881} = 0{,}993\,590.$$

Probabilitatea de deces este

$$q = \frac{1\,044}{162\,881} = 0{,}006\,410.$$

În baza acestor probabilități, deoarece în regiunea A au fost 65 343 locuitori, rezultă că după un an ar fi trebuit în mod teoretic să avem

$$65\,343\,p = 64\,924$$

supraviețuitori și

$$65\,343\,q = 419$$

decedați.

În mod analog, în regiunea B ar fi trebuit să existe în mod teoretic

$$97\,538\,p = 96\,913$$

supraviețuitori și

$$97\,538\,q = 625$$

decedați.

Prin urmare, valorile teoretice corespunzătoare tabelului 100 sînt date de tabelul 102.

Tabelul 102

Regiuni	Numărul teoretic al supraviețuitorilor după un an	Numărul teoretic al decedaților în timp de un an	Total
0	1	2	3
A	64 924	419	65 343
B	96 913	625	97 538
Total:	161 837	1 044	162 881

Valoarea lui χ^2 — deoarece cunoaștem valorile empirice și valorile teoretice corespunzătoare — rezultă imediat

$$\chi^2 = \frac{(64\,904 - 64\,924)^2}{64\,924} + \frac{(96\,933 - 96\,913)^2}{96\,913} + \frac{(439 - 419)^2}{419} + \frac{(605 - 625)^2}{625} = 1{,}535.$$

Efectuînd același calcul pentru toate grupele de vîrstă, obținem:

Grupe de vîrstă	χ^2
15—19	1,535
20—24	1.298
25—29	0,175
30—34	0,047
35—39	0,033
40—44	0,203
45—49	0,420
50—54	1,938
55—59	1,393
Total	7,042

Valoarea totală a lui χ^2 pentru nouă grade de libertate, este prin urmare 7,042. Valoarea lui χ^2 fiind inferioară lui $\chi^2(0{,}05\,;\,9) = 16{,}9$ dată de tabelul anexă II, rezultă că mortalitatea populației din cele două regiuni diferă net nesemnificativ.

EXEMPLUL 3 *). Măsurînd bătaia diametrului 3,4 față de diametrul 5,6 al acului pulverizatorului tractorului KD-35 pentru 249 piese, s-a obținut următoarea distribuție empirică:

Tabelul 103

Bătăile diametrului 3,4 față de diametrul 5,6 în mm (x)	Numărul pieselor (λ_i)
1	2
0,00	7
0,01	60
0,02	95
0,03	57
0,04	19
0,05	6
0,06	3
0,07	2
Total:	249

*) Exemplul este dat în articolul autorilor N. Rancu și L. Tövissi, *Controlul statistic al caracteristicilor strict pozitive ale calității*. Standardizare, 2 (1957).

Distribuției empirice îi corespunde o lege teoretică de distribuție, a cărei densitate este

$$\varphi(x) = \frac{x}{\sigma_0^2} e^{-\frac{x^2}{2\sigma_0^2}}, \qquad \text{(XXII,3.2)}$$

a cărei abatere medie pătratică teoretică este $\sigma = 0{,}655\ \sigma_0$

Cu ajutorul criteriului χ^2 se verifică dacă distribuției celor 249 ace ale pulverizatorului după bătaia diametrului 3,4 față de diametrul 5,6 îi corespunde legea teoretică, dată prin (XXII, 3.2). Pentru aceasta trebuie să se calculeze valoarea lui χ^2 pe baza datelor tabelului 103, în ipoteza că legea de distribuție este (XXII, 3.2).

χ^2 se calculează după formula cunoscută (XXII, 3.1).

Dintre elementele care intervin în această formulă se cunosc frecvențele λ_i care sînt date în coloana 2 a tabelului 103 și mărimea lotului cercetat 249 bucăți. Nu se cunosc însă probabilitățile p_i.

În general:

$$p_i = F(z_i) - F(z_{i-1}) \qquad \text{(XXII, 3.3)}$$

în care:

$$z_i = \frac{x_i}{\sigma_0},$$

$$z_{i-1} = \frac{x_{i-1}}{\sigma_0},$$

unde x_i este limita superioară a intervalului i, iar x_{i-1} limita inferioară a aceluiași interval sau limita superioară a intervalului precedent.

În articolul citat se dau valorile funcției

$$F(z) = \int_0^z \frac{x}{\sigma_0^2} e^{-\frac{x^2}{2\sigma_0^2}} dx$$

pentru z variind între 0 și 4,6.

În exemplul considerat măsurătorile au fost executate cu un aparat de măsurat avînd precizia de citire de 0,01 mm, ceea ce înseamnă că amplitudinea de variație a fost împărțită în intervale a căror mărime este egală cu 0,01 mm și ale căror limite sînt: 0,000; 0,005; 0,015 etc.

Scanned with OKEN Scanner

În tabelul 104 sînt calculate valorile necesare formulei (XXIII, 3.3).

Tabelul 104

x	limitele intervalelor	λ_i	$z_i=\frac{x_i}{\sigma_0}$	$F(z_i)$	p_i	np_i	λ_i-np_i	$(\lambda_i-np_i)^2$	$\frac{(\lambda_i-np_i)^2}{np_i}$
1	2	3	4	5	6	7	8	9	10
0,00	0,000 – 0,005	7	0,27	0,0357	0,0357	8,9	– 1,9	3,61	0,4
0,01	0,005 – 0,015	60	0,82	0,2856	0,2499	62,1	– 2,1	4,41	0,1
0,02	0,015 – 0,025	95	1,37	0,6087	0,3231	81,0	+14,0	196,00	2,4
0,03	0,025 – 0,035	57	1,91	0,8386	0,2299	57,0	0,0	0,00	0,0
0,04	0,035 – 0,045	19	2,46	0,9515	0,1129	28,0	– 9,0	81,00	2,3
0,05	0,045 – 0,055	6	3,04	0,9902	0,0387	9,6	– 3,6	12,96	1,3
0,06	0,055 – 0,065	3 } 5	4,10	0,9997	0,0098	2,4	+ 2,6	6,76	2,8
0,07	0,065 – 0,075	2 }							
	Total	$n=249$	–	–	–	249,0	0,0	–	$\Sigma\frac{(\lambda_i-np_i)^2}{np_i}=9,3$

Valorile lui z_i în coloana 4 a tabelului 104 sînt calculate pe baza relației $z_i = \frac{x_i}{\sigma_0}$, în care x_i sînt limitele superioare ale intervalelor care sînt calculate în coloana 2; σ_0 se obține cu ajutorul relației $\sigma_o = \frac{\sigma}{0,655} = 1,5246\,\sigma$, unde σ abaterea medie pătratică empirică σ_x a cărei valoare este egală, în cazul exemplului cu 0,012.

Se obține astfel:

$$\sigma_0 = 1,524 \times 0,012 = 0,0183.$$

În coloana 5 a tabelului 104 sînt calculate valorile funcției $F(z_i)$ corespunzătoare limitelor superioare ale intervalelor.

În coloana 6 sînt calculate probabilitățile p_i după relația dată mai sus (XXII, 3.3).

Totalul coloanei 10 reprezintă valoarea lui χ^2. Valoarea tabelară χ_p se găsește în tabelul valorilor lui χ^2_p în funcție de probabilitatea p aleasă și de numărul gradelor de libertate K.

Pentru ca un interval să fie luat în considerație, trebuie ca frecvența corespunzătoare să nu fie mai mică decît 5. În cazul cînd frecvențele sînt mai mici decît 5, intervalele respective se contopesc cu alte intervale. Astfel, în cazul nostru, ultimele două intervale au fost contopite într-unul singur.

Ținînd seama de aceasta și de faptul că legea de distribuție depinde de un singur parametru σ_0, numărul gradelor de libertate este:

$$K = 7 - 1 - 1 = 5.$$

Pentru $p = 0,95$ și $K = 5$ găsim $\chi^2_a = 11,1$.

Întrucît $\chi^2 = 9,3$ este mai mic decît $\chi^2_a = 11,1$, ipoteza că legea teoretică de distribuție (XXIII, 3.2) corespunde cu aproximație distribuției empirice se admite ca justă.

4. Distribuția lui Student

Distribuția lui Student, a cărui formulă am dat-o în capitolul VII, are numeroase aplicații statistice. Distribuția teoretică a lui Student este independentă de dispersia legii normale, în raport cu care a fost definită, și ca atare se pot construi funcții de selecție care se distribuie după legea lui Student.

Să considerăm astfel o selecție de mărimea n :

$$x_1, x_2, \ldots, x_n$$

a cărei lege teoretică de probabilitate este $N(m, \sigma)$ adică distribuția normală cu media m și dispersia σ.

Notăm, ca de obicei,

$$\bar{x} = \frac{x_1 + x_2 + \ldots + x_n}{n}; \quad s = \sqrt{\frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 \ldots + (x_n - \bar{x})^2}{n}}.$$

Raportul

$$t = \frac{\sqrt{n-1}\,(\bar{x} - m)}{s}$$

urmează legea lui Student cu $n - 1$ grade de libertate.

Rezultă de aici un criteriu pentru verificarea ipotezei $m = m_0$.

EXEMPLUL 1. Să considerăm următoarea aplicație : într-o întreprindere industrială se fabrică în serie șuruburi cilindrice cu diametrul de 9 mm ; cu ajutorul unor aparate de măsură de precizie, s-au măsurat diametrele a șapte șuruburi luate la întîmplare, constatîndu-se următoarele valori, în milimetri,

$$8{,}75\,;\ 8{,}99\,;\ 9{,}25\,;\ 9{,}32\,;\ 9{,}42\,;\ 9{,}45\,;\ 9{,}50.$$

Se poate admite ipoteza că numerele de mai sus sînt frecvențele unei legi normale, ale cărei medii sînt 9 ?

Avem :

$$\bar{x} = \frac{8{,}75 + 8{,}99 + 9{,}25 + 9{,}32 + 9{,}42 + 9{,}45 + 9{,}50}{7} = 9{,}24,$$

$$s^2 = \frac{\sum_{i=1}^{7} (x_i - \bar{x})^2}{7} = \frac{\Sigma\,(8{,}75 - 9{,}24)^2}{7} = 0{,}064728,$$

$$\mu = \sqrt{\frac{s^2}{6}} = \sqrt{\frac{0{,}064728}{6}} = 0{,}1036.$$

Cu o probabilitate egală cu 0,99 corespunzătoare la șase grade de libertate, trebuie să avem

$$|\,\bar{x} - 9\,| < 3{,}143 \cdot 0{,}1036 = 0{,}33$$

de unde

$$8{,}67 < \bar{x} < 9{,}33.$$

Scanned with OKEN Scanner

Media $\bar{x} = 9{,}24$ este cuprinsă între aceste două limite. Am putea, deci, admite în baza acestui criteriu, cu o probabilitate egală cu 0,99, că media şirurilor de valori considerate, extrase dintr-o distribuţie normală este egală cu 9.

Cu toate că media aparţine intervalului de încredere găsit, totuşi nu putem admite că frecvenţele aparţin unei distribuţii normale.

Într-adevăr, trei frecvenţe (9,42; 9,45; 9,50) ies din intervalul 8,67—9,33, ceea ce nu este verosimil, deoarece probabilitatea ca să obţinem o valoare în afara acestui interval este 0,01. Deci, cu atît mai mult nu putem admite ca din trei probe să se verifice trei evenimente de probabilitate 0,01.

EXEMPLUL 2. Pentru ca un tub fluorescent să fie corespunzător trebuie să aibă o durată de ardere de 2 000 ore. În scopul verificării calităţii loturilor zilnice de tuburi fluorescente din punct de vedere al duratei de ardere se extrag selecţii compuse din 20 de produse. Deoarece nu se cunoaşte abaterea medie pătratică ce caracterizează variaţia duratei de ardere a tuburilor se utilizează abaterea medie pătratică de selecţie.

Rezultatele determinării duratei de ardere a celor 20 tuburi care compun selecţia sînt date în tabelul 105. Se vede că s-a ales origine de calcul valoarea $c = 2\,000$ şi că abaterile duratelor de ardere faţă de durata medie s-au împărţit la 20 adică $a = 20$.

Tabelul 105

Durata de ardere x_i	Numărul tuburilor care au durata x n_i	$x_i' = \frac{x_i - c}{a}$	$n_i x_i'$	$n_i x_i'^2$
1960	2	−2	−4	8
1980	4	−1	−4	4
2000	9	0	0	0
2020	3	1	3	3
2040	2	2	4	8
Total	20		−8 +7 −1	23

Media şi dispersia de selecţie au valorile:

$$\bar{x} = \frac{\Sigma x_i'}{n} 20 + 2\,000 = \frac{-1}{20} 20 + 2\,000 = 1999;$$

$$s^2 = 20^2 \left[\frac{\Sigma x_i'^2}{n} - \left(\frac{\Sigma x_i'}{n} \right)^2 \right] = 20^2 \left[\frac{23}{20} - 0{,}05^2 \right] \approx 460$$

deci abaterea medie pătratică este $s = \sqrt{460}$.

Dacă se notează $m_0 = 2\,000$ ore și cu m durata medie de ardere a lotului de tuburi fluorescente, pentru a verifica ipoteza $m = m_0$ se aplică criteriul:

$$t = \frac{\sqrt{(n-1)}(\bar{x} - m_0)}{s} = \frac{\sqrt{(20-1)}(1999-2000)}{\sqrt{460}} = \frac{\sqrt{19}(-1)}{\sqrt{460}} =$$

$$= \frac{-4 \cdot 36}{21,5} \approx -0,20.$$

Probabilității $p = 0,97$ și numărului gradelor de libertate $n - 1 = 19$ îi corespunde $t = 2,0$.

Deoarece $|-0,20| < 2,0$ se poate afirma cu o probabilitate $p = 0,97$ că durata de ardere medie a tuburilor fluorescente este corespunzătoare.

EXEMPLUL 3. Pe un strung automat cu 6 bare se prelucrează diametrul de 10,2—10,02 mm al pistonașului pompei de injecție a tractorului KD-35. În timpul unei zile, mașina a fost reglată de două ori, adică s-au obținut două loturi de pistonașe corespunzînd celor două reglări.

Pentru a verifica dacă mașina a fost reglată la aceeași dimensiune, din fiecare lot de pistonașe se extrage cîte o selecție compusă din 15 pistonașe.

Rezultatele măsurării diametrului pistonașelor sînt date în tabelul 106.

Tabelul 106

Selecția din lotul I x_{1i}	Selecția din lotul II x_{2i}	$x'_{1i} = \frac{x_{2i} - 10,15}{0,01}$	$x'_{2i} = \frac{x_{2i} - 10,15}{0,01}$	x'^2_{1i}	x'^2_{2i}
10,12	10,14	−3	−1	9	1
10,10	10,18	−5	+3	25	9
10,15	10,20	0	+5	0	25
10,18	10,16	3	1	9	1
10,10	10,14	−5	−1	25	1
10,16	10,11	1	−4	1	16
10,16	10,17	1	2	1	4
10,10	10,17	−5	2	25	4
10,11	10,15	−4	0	16	0
10,12	10,21	−3	6	9	36
10,18	10,15	+3	0	9	0
10,17	10,18	2	3	4	9
10,14	10,20	−1	5	1	25
10,10	10,16	−5	6	25	36
10,16	10,19	1	4	1	16
		−31	− 6	160	183
		11	37		
		−20	+31		

Pe baza tabelului avem :

$$\bar{x}_1 = \frac{-20}{15}\,0{,}01 + 10{,}15 = -0{,}0133 + 10{,}15 = 10{,}1367 \text{ mm},$$

$$\bar{x}_2 = \frac{31}{15}\,0{,}01 + 10{,}15 = 0{,}0207 + 10{,}15 = 10{,}1705 \text{ mm},$$

$$s_1^2 = 0{,}01^2\left[\frac{160}{15} - \left(\frac{-20}{15}\right)^2\right] = 0{,}01^2\,[10{,}6666 - 1{,}3333^2] =$$

$$= 0{,}01^2\,[10{,}6666 - 1{,}7777] = 0{,}08889$$

$$s_2^2 = 0{,}01^2\left[\frac{183}{15} - \left(\frac{31}{15}\right)^2\right] = 0{,}01^2\,[12{,}2222 - (2{,}0667)^2] =$$

$$= 0{,}01^2\,7{,}951 = 0{,}07951.$$

În ipoteza că maşina a lucrat cu aceeaşi precizie, s_1^2 şi s_2^2 sînt estimaţii ale aceleiaşi dispersii generale, deci ele pot fi reunite pentru a obţine o estimaţie a dispersiei generale, adică avem :

$$s = \sqrt{\frac{(n-1)\,s_1^2 + (n-1)\,s_2^2}{n-1+n-1}} = \sqrt{\frac{s_1^2 + s_2^2}{2}}.$$

Făcînd înlocuirile, găsim :

$$s = \sqrt{\frac{0{,}08889 + 0{,}07951}{2}} = 0{,}29.$$

Pentru a verifica dacă maşina a fost reglată pentru aceeaşi dimensiune, trebuie să verificăm dacă cele două medii de selecţie $\bar{x}_1 = 10{,}1367$ mm şi $\bar{x}_2 = 10{,}1705$ mm diferă sau nu semnificativ. În acest scop utilizăm criteriul

$$t = \frac{\bar{x}_1 - \bar{x}_2}{s\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

cu $f = (n_1 - 1) + (n_2 - 1) = 15 - 1 + 15 - 1 = 30 - 2 = 28$ grade de libertate.

În cazul exemplului avem :

$$t = \frac{10{,}1367 - 10{,}1705}{0{,}29\sqrt{\frac{1}{15} + \frac{1}{15}}} = \frac{-0{,}0338}{0{,}29\sqrt{\frac{2}{15}}} \approx -0{,}32.$$

Pentru $p = 0{,}99$ şi $f = 28$ găsim $t_{0{,}99;\,28} = 2{,}467$. Deoarece valoarea calculată $t = 2{,}7$ este mai mare decît valoarea tabelară $t_{0{,}99;\,28} =$

$= 2{,}467$ respingem ipoteza că cele două medii de selecție diferă întîmplător, adică respingem ipoteza că mașina a fost reglată pentru aceeași dimensiune.

EXEMPLUL 4. Fabricarea rezistențelor ohmice după un anumit procedeu tehnologic s-a dovedit a fi oarecum anevoioasă de aceea s-a elaborat un nou procedeu mai rapid. Pentru a decide dacă noua metodă este corespunzătoare din punct de vedere economic și anume dacă este mai eficientă decît prima metodă, s-a efectuat o cercetare selectivă. S-au făcut cîte 10 determinări pentru fiecare procedeu.

Rezultatele se dau în tabelul 107.

Tabelul 107

Rezistențe ohmice obținute prin primul procedeu x_{1i}	Rezistențe ohmice obținute prin al doilea procedeu x_{2i}	$x'_{1i} = \frac{x_{1i}-10{,}5}{0{,}1}$	$x'_{2i} = \frac{x_{2i}-9{,}8}{0{,}1}$	x'^2_{1i}	x'^2_{2i}
10,5	10,1	0	3	0	9
10,2	9,7	− 3	− 1	9	1
10,4	9,3	− 1	− 5	1	25
10,1	9,7	− 4	− 1	16	1
10,8	9,9	3	1	9	1
10,6	9,5	1	− 3	1	9
10,3	9,8	− 2	0	4	0
10,2	9,9	− 3	1	9	1
10,3	9,5	− 2	− 3	4	9
10,6	9,6	1	− 2	1	4
Total		−15	−15		
		+ 5	+ 5	54	60
		−10	−10		

Pe baza tabelului avem:

$$\bar{x}_1 = \frac{-10}{10}0{,}1 + 10{,}5 = -0{,}1 + 10{,}5 = 10{,}4,$$

$$\bar{x}_2 = \frac{-10}{10}0{,}1 + 9{,}8 = -0{,}1 + 9{,}8 = 9{,}7,$$

$$s_1^2 = 0{,}1^2\left[\frac{54}{10} - \left(\frac{-10}{10}\right)^2\right] = 0{,}1^2\ (5{,}4 - 1) = 0{,}1^2 \times 4{,}4 = 0{,}044,$$

$$s_2^2 = 0{,}1^2\left[\frac{60}{10} - \left(\frac{-10}{10}\right)^2\right] = 0{,}1^2\left(\frac{60}{10} - 1\right) = 0{,}1^2 \times 5 = 0{,}05.$$

Scanned with OKEN Scanner

Se vede că cele două dispersii diferă foarte puțin între ele, așa încît pot fi considerate ca estimații ale aceleiași dispersii generale pentru care se calculează estimația

$$s^2 = \frac{n_1 s_1^2 + n_2 s_2^2}{n_1 + n_2 - 2} = \frac{10 \times 0{,}044 + 10 \times 0{,}050}{10 + 10 - 2} = \frac{10}{18} 0{,}094 = 0{,}052$$

de unde

$$s = \sqrt{0{,}052} = 0{,}228.$$

Pentru a decide dacă cea de a doua metodă este sau nu mai eficientă decît prima, aplicăm criteriul

$$t = \frac{\bar{x}_1 - \bar{x}_2}{s\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = \frac{10{,}4 - 9{,}7}{0{,}228\sqrt{\frac{1}{10} + \frac{1}{10}}} = \frac{10{,}4 - 9{,}7}{0{,}228\sqrt{\frac{2}{10}}} = \frac{0{,}7}{0{,}102} = 6{,}9$$

cu $f = 18$ grade de libertate.

În tabele pentru $p = 0{,}98$ și $f = 18$ găsim $t_{0.98:\,18} = 2{,}552$.

Deoarece valoarea obținută prin calcule 6,9 este mai mare decît valoarea tabelară 2,552 iar $\bar{x}_1 > \bar{x}_2$, ipoteza nu se respinge, adică cel de-al doilea procedeu este mai eficient decît primul.

5. Distribuția lui Kolmogorov

Distribuția lui Kolmogorov oferă cel mai însemnat mijloc de verificare pentru ipoteze cu distribuția continuă. Ea face parte din capitolul important al statisticii neparametrice, adică al statisticii în care verificarea statistică se face în mod independent de mărimea unor parametri care caracterizează distribuția.

În statistica neparametrică au un rol deosebit patru criterii cu caracter asimptotic privitoare la abaterile relative și absolute a două funcții empirice de repartiție, respectiv a funcției de repartiție empirică față de cea teoretică.

Trei dintre acestea se datoresc lui Smirnov și unul lui Kolmogorov. Dacă se notează cu $F_m^{(1)}(x)$, $F_n^{(2)}(x)$ funcțiile de repartiție empirică ale variabilelor aleatoare independente X_1 și X_2 corespunzătoare unui număr de m respectiv n probe independente între ele, două din criteriile lui Smirnov se exprimă prin relațiile:

$$\lim_{m;\,n\to\infty} P\left[\sup_{-\infty<x<\infty}\left[F_m^{(1)}(x) - F_n^{(2)}(x)\right] < z\sqrt{\frac{m+n}{mn}}\right] = 1 - e^{-2z^2},$$

$$\lim_{m;\,n\to\infty} P\left[\sup_{-\infty<x<\infty}\left|F_m^{(1)}(x) - F_n^{(2)}(x)\right| < z\sqrt{\frac{m+n}{mn}}\right] = \sum_{k=-\infty}^{+\infty}(-1)^k e^{-2k^2z^2}.$$

De asemenea dacă $F_n(x)$ este funcția de repartiție empirică a valorilor aleatoare mai mici decît x corespunzătoare unui număr de n probe

independente, iar $F(x)$ este funcția de repartiție teoretică, atunci al treilea criteriu al lui Smirnov și criteriul lui Kolmogorov se exprimă prin relațiile:

$$\lim_{n\to\infty} P\left[\sup_{-\infty<x<\infty}[F_n(x) - F(x)] < \frac{z}{\sqrt{n}}\right] = 1 - e^{-2z^2},$$

$$\lim_{n\to\infty} P\left[\sup_{-\infty<x<\infty}|F_n(x) - F(x)| < \frac{z}{\sqrt{n}}\right] = \sum_{k=-\infty}^{+\infty} (-1)^k e^{-2k^2z^2}.$$

O simplă privire atrage atenția asupra legăturii strînse dintre expresiile ce s-au obținut pentru abaterile relative și cele absolute și anume la cele dintîi apare cîte o exponențială, la celelalte, sume algebrice de exponențiale.

Se pune în mod logic problema de a găsi prin mijloace elementare expresia asimptotică pentru abateri absolute, folosind expresia pentru abateri relative, cu atît mai mult cu cît pentru abateri relative problema se rezolvă incomparabil mai simplu.

Astfel, la al treilea criteriu al lui Smirnov se poate ajunge pe o cale cu totul elementară dacă se folosește expresia exactă pentru

$$P[\sup_{-\infty<x<\infty}[F_n(x) - F(x)] < y] = P[\sup_{-\infty<x<\infty}[F(x) - F_n(x)] < y].$$

La aceasta se poate ajunge de exemplu cum a făcut Koroliuk *) pornind de la expresia exactă $P[\sup [F^{(1)}_{np}(x) - F_n(x) < y]$ care poate fi calculată cu ușurință prin mijloace elementare și făcînd $p \to \infty$

Se obține astfel

$$P[\sup_{-\infty<x<\infty}[F(x) - F_n(x)] < y] = 1 - \sum_{i=0}^{[n(1-y)]} \binom{n}{i} y\left(y + \frac{i}{n}\right)^{i-1}\left(1 - y - \frac{i}{n}\right)^{n-i}.$$

Făcînd $y = \frac{z}{\sqrt{n}}$ și $n \to \infty$ se obține cu ajutorul unor calcule laborioase, dar elementare, expresia asimptotică $1 - e^{-2z^2}$ al celui de-al treilea criteriu al lui Smirnov.

Nu expunem demonstrația lui Koroliuk. Vom trasa însă în linii largi demonstrația pe care G. Theiler **) a dat-o teoremei lui Kolmogorov, considerînd că aceasta constituie, prin ideile pe care le conține, o contribuție însemnată la teoria statisticii neparametrice.

G. Theiler a stabilit o lemă care permite să se obțină foarte ușor expresiile corespunzătoare criteriilor lui Kolmogorov și Smirnov pentru abateri absolute folosind expresiile pentru abateri relative.

În acest scop s-a pornit de la faptul că orice problemă privitoare la abaterile a două funcții empirice de repartiție, ca mai sus, este echivalentă

*) *O rashojdenii empiriceskih raspredelenii dlea sluciaia dvuh nezavisimih viborok.* Izvestia Akad. nauk S.S.S.R., **19**, 369–397 (1955).

**) *Asupra criteriilor lui Kolmogorov și Smirnov din teoria statisticei neparametrice.* Comunicările Academiei R.P.R., *8* (1961).

cu problema corespunzătoare privind extragerea pînă la epuizare a bilelor dintr-o urnă care conține m bile albe și n bile negre, identice între ele. Se poate arăta că $\sup [F_m^{(1)}(x) - F_n^{(2)}(x)] = \max [S_k - T_k]$, unde $S_k = \frac{X_k}{k}$; $T_k = \frac{Y_k}{k}$ iar X_k și Y_k reprezintă numărul bilelor albe și negre ce se găsesc între primele k bile dintr-o extragere pînă la epuizare a bilelor din urnă.

Se poate arăta apoi că:

$$P\left[\sup_{-\infty<x<\infty} |F_m^{(1)}(x) - F_n^{(2)}(x)| < y\right] = 1 + \sum_{i+j=1} (-1)^{i+j} [N_{i,j} + N'_{i,j}] : \binom{N}{n}, \quad \text{(XXII, 5.1)}$$

unde

$$j = i - 1\,;\ i\,;\ i + 1$$

iar

$$N_{i,j} = \Sigma N [E_{k_{r_1}; k_{r_2}; \ldots; k_{r_{i+j}}}]\,;\ N'_{i,j} = \Sigma N [E_{k_{r_0}; k_{r_1}; \ldots; k_{r_{i+j-1}}}]$$

prin $N[\ldots]$ notîndu-se numărul total al grupelor de $(m + n)$ bile în care au loc evenimente de tipul celor descrise în paranteză.

Pentru a arăta ce reprezintă evenimentele descrise în paranteze trebuie introduse în prealabil următoarele notații:

$m = m_1 d$; $n = n_1 d$, unde d este cel mai mare divizor comun al numerelor m și n;

$y_1 = -[-m_1 n_1 d y]$, prin $[\ldots]$ înțelegînd cel mai mare întreg conținut în paranteză;

$$k_{r_{2u+1}} = \left[\frac{(m_1 + n_1) r_{2u+1} + y_1 + n_1 - 1}{n_1}\right] \text{ cu } 0 \leqslant r_{2u+1} \leqslant n - \left[\frac{y_1}{m_1}\right]$$

$$k_{r_{2v}} = \left[\frac{(m_1 + n_1) r_{2v} - y_1}{n_1}\right] \text{ cu } \left[\frac{y_1 + m_1 - 1}{m_1}\right] \leqslant r_{2v} \leqslant n.$$

Acum putem preciza: prin $E_{k_{r_1}; k_{r_2}; \ldots; k_{r_{i+j}}}$ se înțelege o grupă de $(m + n)$ bile dintre care m albe și n negre, — în care r_1 este cel mai mic număr pentru care în grupă există un rang $k_{r_{2u+1}}$ pînă la care (inclusiv) numărul bilelor negre să fie r_{2u+1}; r_2 cel mai mic număr r_{2v} pentru care în grupă să existe după rangul k_{r_1} un rang $k_{r_{2v}}$, pînă la care (inclusiv) să existe r_{2v} bile negre ș.a.m.d.

În mod analog se definește $E_{k_{r_0}; k_{r_1}; \ldots; k_{r_{i+j-1}}}$.

Notațiile de mai sus servesc în principal la aflarea expresiilor exacte pentru unele probleme de statistică neparametrică, dar ele pot fi folosite și pentru aducerea problemelor de statistică neparametrică cu caracter asimptotic privind abaterile absolute la cele privitoare la abateri relative, pe baza următoarei leme:

Scanned with OKEN Scanner

LEMĂ. Este valabilă dubla relație:

$$P\left[\sup_{-\infty<x<\infty}[F_m^{(1)}(x)-F_n^{(2)}(x)]\geqslant\frac{(i+j)\,y_1}{m\,n_1}\right]>\frac{N_{i;j}}{\binom{N}{n}},$$

respectiv

$$\frac{N'_{i;j}}{\binom{N}{n}}>P\left[\sup_{-\infty<x<\infty}[F_m^{(1)}(x)-F_n^{(2)}(x)]\geqslant\left[\frac{(i+j)\,y_1+n_1\,i+m_1 j}{m\,n_1}\right]\right].$$

Demonstrația acestei leme are un caracter elementar. Dacă se ia $y=\frac{z}{\sqrt{n}}$; $m=np$, deci $n_1=1$; $m_1=p$ și se face $p\to\infty$ se obține:

$$P\left[\sup_{-\infty<x<\infty}[F(x)-F_n(x)]\geqslant\frac{(i+j)\,z}{\sqrt{n}}\right]\geqslant\frac{N_{i,j}^{(1)}}{n^n};$$

$$\frac{N_{i,j}^{(1)}}{n^n}\geqslant P\left[\sup_{-\infty<x<\infty}[F(x)-F_n(x)]\geqslant\frac{(i+j)\,z}{\sqrt{n}}+\frac{j}{n}\right],$$

unde

$$\frac{N_{i,j}^{(1)}}{n^n}=\lim_{p\to\infty}\frac{N_{i,j}}{\binom{N}{n}}\text{ și }\frac{N_{i,j}^{\prime(1)}}{n^n}=\lim_{p\to\infty}\frac{N'_{i,j}}{\binom{N}{n}};$$

aceste limite existînd și expresiile lor fiind binecunoscute.

Dînd lui i și j valori determinate și făcînd $n\to\infty$ se obține:

$$\lim_{n\to\infty}\frac{N_{i,j}^{(1)}}{n^n}=\lim_{n\to\infty}\frac{N_{i,j}^{\prime(1)}}{n^n}=e^{-2(i+j)^2z^2}.$$

Pornindu-se apoi de la expresia exactă:

$$P\left[\sup|F(x)-F_n(x)|<\frac{z}{\sqrt{n}}\right]=1+\sum_{i+j=1}(-1)^{i+j}\left[\frac{N_{i,j}^{(1)}}{n^n}+\frac{N_{i,j}^{\prime(1)}}{n^n}\right]$$

(XXII, 5.2)

care se deduce din expresia generală (XXII, 5.1) făcînd $y=\frac{z}{\sqrt{n}}$ și $m\to\infty$ (cel mai simplu luînd $m=np$ și făcînd $p\to\infty$), se obține imediat expresia dată de Kolmogorov.

Într-adevăr, observăm, că înșiși prin natura lor $N_{i+j}^{(1)}$ respectiv $N_{i+j}^{\prime(1)}$ descresc cînd $i+j$ crește, iar în (XXII, 5.2) avem o serie alternată cu termeni descrescători, restul fiind deci în valoare absolută mai mic ca ultimul termen și tinzînd deci la zero cînd $(i+j)\to\infty$.

Putem deci scrie că limita sumei algebrice este egală cu suma limitelor, ceea ce epuizează demonstrația.

Vom aplica acum criteriul lui Kolmogorov la un exemplu.

EXEMPLUL 1. S-a măsurat o caracteristică tehnică a 150 piese de tractor și rezultatele au fost trecute în tabelul 108. Ne propunem să verificăm prin criteriul lui Kolmogorov normalitatea acestor date statistice.

În tabelul 108 s-a calculat media $\bar{x} = 15{,}55$ și dispersia $s = 0{,}02629$.

Tabelul 108

Nr. crt.	x_i	f_i	$x_i' = \frac{x_i - c}{a}$	$x_i' f_i$	$x_i'^2 f_i$
1	15,50	6	−5	−30	150
2	15,51	6	−4	−24	96
3	15,52	18	−3	−54	162
4	15,53	15	−2	−30	60
5	15,54	15	−1	−15	15
6	15,55	21	0	0	0
7	15,56	19	1	15	19
8	15,57	17	2	34	68
9	15,58	16	3	48	144
10	15,59	11	4	44	176
11	15,60	6	5	30	150
Total		150	—	22	1 040

$$a = 0{,}01,\ c = 15{,}55,$$

$$\bar{x} = 15{,}55\,;\ s = 0{,}02629.$$

În tabelul 109 s-au calculat valorile funcției de repartiție a distribuției normale $F(x_i)$ în punctele x_i. De asemenea au fost calculate valorile funcției de repartiție empirice $F_n(x_i)$ în punctele x_i.

Tabelul 109

x_i	f_i	$F(x_i)$	$F_n(x_i)$	$\lvert F_n(x_i) - F(x_i)\rvert$
15,50	6	0,04363	0,04000	0,00363
15,51	6	0,09176	0,08000	0,01176
15,52	18	0,17106	0,20000	0,02894
15,53	15	0,28434	0,30000	0,01566
15,54	15	0,42465	0,40000	0,02465
15,55	21	0,57535	0,54000	0,03535
15,56	19	0,71566	0,66666	0,04900
15,57	17	0,82894	0,77999	0,04895
15,58	16	0,90824	0,88665	0,02159
15,59	11	0,95637	0,95998	0,00361
15,60	6	0,98169	0,99998	0,01829

Scanned with OKEN Scanner

Criteriul lui Kolmogorov ne spune că

$$P\left[\max | F(x) - F_n(x) | < \frac{\lambda}{\sqrt{n}}\right] = k(\lambda),$$

unde valoarea lui $k(\lambda)$ se ia din tabele.

Pentru n egal cu 150 și $k(\lambda)$ egal cu 0,98, găsim că λ este egal cu 1,5175 și prin urmare, cu o probabilitate egală cu 0,98, trebuie să avem :

$$\max | F(x_i) - F_n(x_i) | < \frac{\lambda}{\sqrt{n}} = \frac{1,5175}{\sqrt{150}} = 0,1239.$$

Din ultima coloană se vede că maximum diferenței absolute dintre cele două funcții este $0,049 < 0,1239$. Prin urmare criteriul este verificat.

EXEMPLUL 2*). Măsurînd gramajul hîrtiei de ziar la 224 determinări s-a obținut distribuția empirică din tabelul 110.

Tabelul 110

Valorile gramajului (g/m^2)	Numărul determinărilor
x_i	f_i
48,5	21
49,0	52
49,5	90
50,0	50
50,5	11
Total :	224

Se poate considera că acestei distribuții empirice îi corespunde legea normală, deoarece nu există nici un motiv a nu admite că erorile de producție sînt generate ca o sumă de un număr foarte mare de erori mici. Se pune deci problema verificării normalității distribuției empirice. Pentru aceasta se folosește criteriul lui Kolmogorov.

Calculele necesare pentru obținerea criteriului lui Kolmogorov sînt date în tabelul 111.

Tabelul 111

Valorile caracteristicii	Limitele intervalelor	Frecvența f_i	$\frac{x_i - \bar{x}}{6} = z_i$	$\Phi(z_i)$	$F(x_i) = 0,5 + \Phi(z_i)$	Frecvența relativă f_r	$F_n(x_i)$	$\lvert F_n(x_i) - F(x_i) \rvert$
48,5	48,25—48,75	21	−1,4	−0,419	0,081	0,093	0,093	0,012
49,0	48,75—49,25	52	−0,4	−0,155	0,345	0,232	0,325	0,020
49,5	49,25—49,75	90	+0,6	0,226	0,726	0,402	0,725	0,001
50,0	49,75—50,25	50	+1,6	0,445	0,945	0,223	0,950	0,005
50,5	50,25—50,75	11	+2,6	0,495	0,995	0,049	0,999	0,004

*) Exemplul este dat de N. Rancu și L. Tövissi, *Aplicații ale statisticii în industria hîrtiei*. Celuloză și Hîrtie, *1* (1958).

În tabelul 111 s-a notat cu x_i limita superioară a intervalului i, cu $\Phi(z_i)$ funcția lui Laplace, ale cărei valori corespunzătoare unui z_i se găsesc în tabelul anexă 1. În calcul s-au folosit pentru $\bar{x}$ și s valorile obținute pe baza datelor din tabelul 111, și anume: $\bar{x} = 49{,}45$, respectiv $s = 0{,}05$. Din acest tabel rezultă că diferența maximă este 0,020, dar D_n este valoarea absolută a diferenței maxime, adică 0,020. Alegînd pentru q_λ valoarea $q_\lambda = 0{,}01$, se găsește că $\lambda = 1{,}63$. Știind că numărul încercărilor este 224, se obține $\frac{\lambda}{\sqrt{n}} = \frac{1{,}63}{\sqrt{224}} = 0{,}108$. Comparînd valoarea lui $D_n = 0{,}020$ cu 0,108 se conchide potrivit criteriului lui Kolmogorov că distribuției empirice îi corespunde legea normală.

Același rezultat se obține și aplicînd criteriul χ^2.

Mărimea

$$\chi^2 = \Sigma \frac{(f_i - np_i)^2}{np_i}$$

este calculată în tabelul 112.

Tabelul 112

Valorile caracteristicii	Limitele intervalelor	f_i	$z_i = \frac{x_i - \bar{x}}{\sigma}$	$\Phi(z_i)$	$p_i = \Phi(z_i) - \Phi(z_{i-1})$	np_i	$f_i - np_i$	$(n - np_i)^2$	$\frac{(n - np_i)^2}{np_i}$
48,5	48,25—48,75	21	−1,4	−0,419	0,081	18	3	9	0,50
49,0	48,75—49,25	52	−0,4	−0,155	0,264	59	−7	49	0,83
49,5	49,25—49,75	90	+0,6	0,226	0,380	85	5	25	0,30
50,0	49,75—50,25	50	+1,6	0,445	0,219	49	1	1	0,02
50,5	50,25—50,75	11	+2,6	0,495	0,055	12	−1	1	0,09
Total	—	—	—	—	—	—	—	—	$\chi^2 = 1{,}74$

Valoarea lui χ^2 se obține prin însumarea valorilor din ultima coloană a tabelului 112. Această valoare se compară cu valoarea lui χ^2_q corespunzătoare unui q dat. Valoarea lui q se alege cît mai aproape de zero și reprezintă probabilitatea ca χ^2 să fie mai mare decît χ^2_q. Valoarea lui χ^2 se găsește în tabelul anexă 2, corespunzătoare numărului gradelor de libertate k și valorii alese pentru q.

După cum știm, numărul gradelor de libertate se calculează cu ajutorul formulei $k = N - I - 1$, unde N este numărul intervalelor, iar I — numărul parametrilor legii de distribuție. În cazul exemplului considerat, $N = 5$, iar $I = 2$, deoarece legea normală are doi parametri: $\bar{x}$ și σ.

Alegînd pentru q valoarea 0,1, iar k fiind egal cu 2, se găsește în tabel pentru χ^2_q valoarea 4,6. Comparînd valoarea lui $\chi^2 = 1{,}74$ cu valoarea lui $\chi^2_q = 4{,}6$ se conchide că distribuției empirice îi corespunde legea normală, adică gramajul hîrtiei de ziar se distribuie după legea normală.

Scanned with OKEN Scanner

EXEMPLUL 3*). Cercetînd numărul ruperilor firului la o mașină cu inele cu 224 fuse, timp de 8 ore s-a obținut distribuția din tabelul de mai jos.

Tabelul 113

Numărul ruperilor x	Numărul fuselor la care firul s-a rupt de x ori n_x	Numărul total al ruperilor $x\,n_x$
0	92	0
1	86	86
2	34	68
3	9	27
4	2	8
5	1	5
Total	224	194

Firul se poate rupe la fiecare rotire a cursorului care execută în curs de 8 ore un număr de rotații de ordinul milioanelor, deci ruperea firului este un eveniment rar, de aceea se poate face ipoteza că distribuția numărului de ruperi ale firului urmează legea lui Poisson a cărei funcție de frecvență este

$$f(x) = \frac{a^x e^{-a}}{x!}.$$

Ținînd seamă că

$$a = M(x) = \frac{194}{224} = 0{,}866,$$

obținem

$$f(0) = e^{-0.866} = 0{,}42063.$$

Celelalte valori ale funcției de frecvență se calculează utilizînd formula de recurență

$$f(x+1) = f(x)\frac{a}{x+1}$$

și sînt date în tabelul 114.

Probabilității $P = k(\lambda) = 0{,}95$ îi corespunde $\lambda = 1{,}36$. Cum numărul observațiilor este $n = 224$, găsim

$$\frac{\lambda}{\sqrt{n}} = \frac{1{,}36}{\sqrt{224}} = 0{,}0909.$$

Din tabelul 114 (ultima coloană) se vede că diferența cea mai mare 0,00992 este mai mică decît 0,0909.

*) N. Rancu și L. Tövissi, *Metode statistico-matematice de control al numărului de ruperi ale firului*, Industria textilă, 6 (1960).

Tabelul 114

Nr. ruperilor x	$f(x)$	$F(x) = P(X \leqslant x)$	Frecvenţe observate (n_x)	Frecvenţe relative	Frecvenţe relative cumulate $Fn(x)$	$\lvert F(x) - F_n(x) \rvert$
0	0,42063	0,42063	92	0,41071	0,41071	0,00992
1	0,36427	0,78490	86	0,38393	0,79464	0,00974
2	0,15773	0,94263	34	0,15179	0,94643	0,00380
3	0,04553	0,98816	9	0,04018	0,98661	0,00155
4	0,00986	0,99802	2	0,00893	0,99554	0,00248
5	0,00171	0,99973	1	0,00446	1,00000	0,00027
6	0,00025	0,99998	0	0,00000	1,00000	0,00002
7	0,00002	1,00000	0	—	1,00000	0,00000
Total	1,00000		224			

Rezultă că ipoteza făcută este justă, adică se poate afirma cu o probabilitate $p = 0{,}95$ că „numărul de ruperi ale firului" urmează legea de distribuţie a lui Poisson.

Vom pune în evidenţă importanţa practică deosebită a problemei studiate. Dacă examinăm evenimentul $X > x$, observăm că pentru $x = 3$ avem $P(X > 3) = 1 - F(3) = 1 - 0{,}98816 = 0{,}01184$.

În condiţii de producţie se consideră că un eveniment este practic imposibil dacă are o probabilitate mai mică decît 0,02. De aceea se poate afirma că, în condiţiile de producţie date, practic, este imposibil ca numărul de ruperi ale firului la oricare fus să fie mai mare decît 3. Prin urmare, ori de cîte ori se constată că firul s-a rupt de 4 sau de mai multe ori se poate considera că procesul de producţie nu se desfăşoară normal, că a intervenit o cauză care a dus la mărirea numărului de ruperi.

EXEMPLUL 4*). Cercetînd bătaia conului lucrător faţă de gaura 4,93 a corpului pulverizator la un număr de 273 piese s-au obţinut rezultatele de mai jos (tabelul 115).

Tabelul 115

Bătaia x	Numărul pieselor care au bătaia x r_x
0,001	4
0,002	15
0,003	49
0,004	60
0,005	54
0,006	44
0,007	25
0,008	13
0,009	5
0,010	1
0,011	1
0,013	2

*) N. Rancu şi L. Tövissi, *Controlul statistic al caracteristicilor strict pozitive prin metoda probelor ordonate.* Standardizarea, *8* (1957).

Vom verifica dacă distribuția „bătăii conului lucrător față de diametrul 4,93" urmează legea

$$f(x) = \sqrt{\frac{2}{\pi}} \frac{x^2}{\sigma_0^3} e^{-\frac{x^2}{2\sigma_0^2}}$$

unde $\sigma_0 = 1{,}484 \cdot \sigma$.

Calculînd abaterea medie pătratică pe baza celor 273 observații obținem $\sigma = 0{,}00186$.

În continuare, vom calcula valorile funcției de distribuție în ipoteza că variabila cercetată „bătaia conului lucrător față de diametrul 4,93" urmează legea de mai sus, aceste valori se pot lua din tabelul întocmit pentru această lege *).

În tabelul 116 (coloana 5) se dau valorile funcției de distribuție, corespunzătoare limitelor superioare ale intervalelor.

Tabelul 116

Bătaia x	Limitele intervalelor $x_{i-1} - x_i$	Nr. pieselor n_x	$z_i = \frac{x_i}{\sigma_0}$	$F(z_i)$	$nF(z_i)$	Frecvența cumulată (f_c)	$\lvert nF(z_i) - f_c \rvert$
0,001	0,0005—0,0015	4	0,54	0,03839	10,48	4	6,48
0,002	0,0015—0,0025	15	0,90	0,14290	39,00	19	20
0,003	0,0025—0,0035	49	1,27	0,34356	93,70	68	25,7
0,004	0,0035—0,0045	60	1,63	0,55222	150,40	128	22,4
0,005	0,0045—0,0055	54	1,99	0,73418	200,00	182	18
0,006	0,0055—0,0065	44	2,35	0,86268	235,00	226	9
0,007	0,0065—0,0075	25	2,71	0,93832	256,00	251	5
0,008	0,0075—0,0085	13	3,08	0,97652	267,00	264	3
0,009	0,0085—0,0095	5	3,44	0,99202	271,00	269	2
0,010	0,0095—0,0105	1	3,80	0,99769	271,00	270	1
0,011	0,0105—0,0115	1	4,16	0,99939	272,00	271	1
0,013	0,0125—0,0135	2	4,89	1,00000	273,00	273	0
Total		273					

După cum se vede, în acest caz se calculează diferența dintre frecvențele absolute teoretice $nF(z)$ și frecvențele absolute cumulate f_c, adică vom examina evenimentul

$$n.\max \lvert F(x) - F_n(x) \rvert > n \frac{\lambda}{\sqrt{n}}$$

sau

$$\max \lvert nF(x) - f_c \rvert > \sqrt{n}\, \lambda$$

în locul evenimentului

$$\max \lvert F(x) - F_n(x) \rvert > \frac{\lambda}{\sqrt{n}};$$

această înlocuire aduce o importantă simplificare a calculelor.

*) N. Rancu și L. Tövissi, *Controlul statistic al producției*. I.D.T., București, 1958, tabela VI, p. 166.

Pentru probabilitatea $P = k(\lambda) = 0{,}99$ găsim în tabele $\lambda = 1{,}63$ deci $\sqrt{n}\,\lambda = \sqrt{273} \cdot 1{,}63 = 27{,}06$.

Examinînd ultima coloană a tabelului 116, vedem că diferența cea mai mare este $d = 25{,}7$; aceasta este mai mică decît 27,06 deci ipoteza făcută este justă, adică putem afirma cu o probabilitate $P = 0{,}99$ că distribuția „bătăii conului lucrător față de diametrul 4,93" urmează legea presupusă.

6. Extinderea rezultatelor cercetării selecției asupra colectivității generale

Am văzut că prin metoda selecției se pune problema de a fi estimate unele valori tipice ale caracteristicilor necunoscute din colectivitatea generală pe baza datelor obținute prin observarea doar a unei părți din această colectivitate.

Extinderea rezultatelor cercetării selective asupra întregii colectivități se poate face prin două metode:

a) metoda extinderii directe;

b) metoda coeficientului de corecție.

Extinderea directă a datelor selecției asupra colectivității generale se face în mod necesar, cînd nu dispunem decît de aceste date.

În acest scop, pentru fiecare caracteristică din programul observării se consideră nivelul ei mediu, eroarea limită de selecție și volumul colectivității generale.

Să considerăm următorul exemplu: să presupunem că într-un oraș oarecare cu un număr de 100 000 capi de familie se face un sondaj, de 10%, pentru a se stabili volumul economiilor (x_i) la C.E.C. al tuturor familiilor.

Datele privind această selecție sînt următoarele: ($n = 10\,000$); volumul total al economiilor la cele 10 000 de familii cercetate este de 10 000 000 lei

Avem:

s^2 dispersia, egală cu 300 lei.

$\frac{n}{N} = f = 0{,}1$ și $\frac{1}{f} = q = 10$ reprezintă coeficientul de extindere.

Notînd cu X_i' volumul total al economiilor estimat în colectivitatea generală, vom avea

$$X_i' = qx_i = 10 \cdot 10\,000\,000 = 100\,000\,000 \text{ lei.}$$

Eroarea de selecție în acest caz poate fi determinată astfel

$$S_{x_i'} = \sqrt{\frac{ns^2}{f^2}(1 - f)} \qquad \text{(XXII, 6.1)}$$

întrucît se poate demonstra că la estimarea indicatorului general, dispersia este de N^2 ori mai mare decît dispersia medie de selecție corespunzătoare, selecția fiind nerepetată.

Aplicînd în exemplul nostru formula (XXII, 6.1), vom avea

$$S_{x_i} = \sqrt{\frac{10\,000.300}{0{,}01}(1 - 0{,}10)} = \sqrt{300\,000\,000 \times 0{,}90} = 16\,431.$$

Limitele de siguranță cu probabilitatea de 98,95% vor fi deci

$$100\,000\,000 \pm 2{,}56 \cdot 16\,431,$$

adică

$$100\,000\,000 \pm 42{,}063 = \begin{cases} 100\,042\,063 \text{ lei (maximum)}, \\ 99\,957\,937 \text{ lei (minimum)}. \end{cases}$$

Prin urmare, cu o probabilitate de 98,95% se poate afirma că volumul total al economiilor la C.E.C. realizat de toate familiile din orașul nostru este cuprins între 99 957 937 și 100 042 063 lei.

Dacă în exemplul luat de noi am fi avut date de selecție caracterizînd valoarea medie a soldului depunerii la C.E.C. pe care o are un cap de familie, atunci, evident, pentru a afla volumul total al economiilor ar fi trebuit să înmulțim această valoare medie cu 100 000, determinînd totodată și gradul de siguranță probabilistic cu care obținem indicatorul respectiv.

Așa-zisa metodă a coeficientului de corecție se folosește de obicei cînd cercetarea selectivă s-a făcut cu scopul exclusiv de a fi examinate datele culese printr-o înregistrare totală a colectivității generale. E evident că în acest caz nu se lucrează cu nivelul mediu al caracteristicilor din programul observării, ci cu nivelul lor absolut pe colectivitatea cercetată.

Pentru a fi rectificate datele înregistrării totale, ele se înmulțesc cu un coeficient de corecție stabilit printr-un raport în care, la numărător, figurează datele înregistrării de selecție (referitoare la anumite fenomene studiate), iar la numitor, datele înregistrării totale *).

7. Cîteva observații și concluzii generale asupra metodei selecției

Expunerea de pînă acum n-a epuizat, desigur, toate problemele teoretice și practice ale metodei selecției; socotim totuși că au fost abordate cele mai importante teoreme de statistică-matematică pe baza cărora se justifică aplicarea metodei selecției, așa cum o întîlnim în practica statistică obișnuită. Mai mult, au fost abordate chiar și unele scheme teoretice, însoțite de exemplificări practice, care de obicei nu figurează în tratatele de statistică teoretică. De exemplu, au fost examinate problemele teoretice și practice ale selecției prin care se supun studiului nu o singură caracteristică, ci două și chiar mai multe caracteristici, fiecare

*) Vezi un exemplu concret de calcul în lucrarea lui M. Biji și I. Stoichiță, *Metoda selectivă în cercetarea statistică*. Ed. științifică, București, 1957, p. 152—153.

variind într-un anumit interval. Cazurile cînd există mai multe caracteristici se pot considera, de altfel, ca fiind cele mai frecvente în viața social-economică.

În general a fost examinat numai procedeul de determinare prin metoda selecției, a mediei aritmetice a unei caracteristici dintr-o colectivitate A. Aceeași problemă se pune însă și pentru alte valori tipice ale acestei colectivități. Așa ne putem pune problema de a determina valoarea medianei, a modului, a dispersiei unităților care compun colectivitatea generală etc., folosind tot metoda selecției.

Se vede că pentru aplicarea metodei selecției mai sînt încă numeroase probleme teoretice ce se impun a fi rezolvate.

Metoda selecției trebuie să se aplice după un plan bine organizat, conceput de statisticieni-economiști, specializați în problemele de selecție. Cunoștințele de statistică-matematică sînt indispensabile însă nu sînt suficiente.

Cei care concep planul selecției trebuie să fie familiarizați cu dificultățile practice întîlnite în aplicarea anchetelor statistice. Dar nici aceasta nu este suficient.

După cum am văzut, planul de selecție trebuie ales astfel încît să conducă la rezultatele cele mai precise, cu cheltuieli minime. Colectivitatea generală trebuie despărțită în subcolectivități sau grupe tipice astfel determinate încît în interiorul fiecărei grupe atributul caracteristicii supuse studiului să se repartizeze cît mai omogen posibil. Nu există reguli precise pentru determinarea grupelor, există însă principiile generale elaborate de statistica socialistă de care trebuie să se țină seamă pentru a asigura grupării un caracter științific. Se știe că la baza grupării trebuie luate caracteristicile tipice, esențiale ale fenomenului studiat, în conformitate cu sarcinile concrete ce revin cercetării selective. Alegerea caracteristicilor de grupare reclamă analiza multilaterală a fenomenelor social-economice supuse studiului, pe baza teoriei marxist-leniniste, clarificîndu-se, în primul rînd, esența lor și particularitățile lor distincte. Aici intervine rolul cel mai important al celui care elaborează planul de selecție, care, bazîndu-se pe principiile de mai sus, pe experiența lui și pe cunoașterea amănunțită a domeniului în care se face cercetarea selectivă, trebuie să facă cea mai justă grupare a colectivității generale.

O dată grupele stabilite, trebuie să se fixeze numărul unităților care vor fi extrase din fiecare grupă, în raport cu gradul de omogeneitate al grupei, ceea ce implică de asemenea o experiență deosebită din partea celui care întocmește planul lucrării.

După cum am văzut, numerele n_i de extracții din fiecare grupă depind de dispersia valorilor din fiecare grupă. Această dispersie este însă necunoscută. Este adevărat că din examinarea fenomenelor stabilite pînă acum rezultă că pentru aplicarea metodei selecției nu avem nevoie de valoarea exactă a dispersiei, ci numai de o apreciere aproximativă a ei. Această apreciere însă nu o poate face decît un bun cunoscător al domeniului cercetat pe baza unei anchete prealabile de mai mică întindere, ca și a diferitelor studii de specialitate etc. Trebuie să subliniem că aceasta este dificultatea principală în aplicarea metodei selecției, depă-

șind cu mult dificultățile legate de aplicarea tehnicii culegerii materialului statistic sau a formulelor matematice care variază foarte puțin de la un plan la altul și pot fi ușor înlăturate prin aplicarea, aproape automată, a unor reguli fixe, existente în toate tratatele de specialitate.

Folosită judicios, cu respectarea riguroasă a principiilor grupării leniniste, a principiului selecției, metoda selectivă oferă statisticii socialiste din țara noastră, posibilitatea de a îndeplini cu succes sarcinile noi ce-i stau în față — mai ales pe linia ridicării necontenite a calității analizelor statistice — sarcini legate de traducerea în viață a hotărîrilor istorice ale celui de-al III-a Congres al P.M.R.

Scanned with OKEN Scanner

PARTEA
A ȘASEA

METODE
MODERNE
DE CALCUL
STATISTIC

Scanned with OKEN Scanner

Scanned with OKEN Scanner

Capitolul XXIII

Metode statistice de calcul numeric. Calculul cu numere întîmplătoare

1. Considerații generale

Aplicațiile din ce în ce mai numeroase ale statisticii matematice în științele naturii și sociale, conduc la calcule numerice laborioase, pentru a căror executare se cere nu numai precizie, dar și rapiditate. În cuprinsul acestui tratat, am întîlnit probleme a căror rezolvare depinde de calculul unor integrale, de soluționarea unor ecuații diferențiale, de rezolvarea unui sistem de ecuații algebrice liniare, de maximizarea unei funcții de eficiență etc.

La aplicarea metodei programării liniare în economia socialistă vom vedea că pentru determinarea legăturilor între ramuri, este necesară inversarea unei matrici fundamentale. Deoarece această matrice are în general un număr mare de elemente, calculul direct al matricei inverse conduce la calcule foarte complicate. Astfel, dacă matricea are n linii și n coloane pentru determinarea matricei inverse este necesar să se calculeze un determinant de ordinul n și n^2 determinanți de ordinul $n-1$, ceea ce după cum se știe necesită un număr imens de operații, de multe ori greu de executat chiar la mașinile electronice de calcul.

Renunțarea la calculul direct și folosirea în acest scop a metodelor statistice care conduc la soluții cu mare grad de exactitate într-un timp scurt este de aceea foarte folositoare.

Calculul diferitelor valori prin metode statistice este cunoscut de multă vreme. El se bazează pe folosirea unor funcții de selecție și pe convergența lor către anumite limite. Contribuția modernă care a fost adusă acestei metode, constă în folosirea „numerelor întîmplătoare". Numerele întîmplătoare reprezintă consemnarea unor experiențe făcute. Istoria

experienţelor realizate ne dispensează de realizarea unor mari experienţe. Calculele numerice bazate pe metode statistice pretind un număr mare de experienţe, după cum vom vedea pentru calculul valorii lui π. S-au făcut mii de experienţe constînd în aruncarea unui ac sau au fost aruncate zaruri de mii de ori. În prezent nu mai avem nevoie să facem asemenea experienţe, deoarece folosim rezultatele unor experienţe anterioare, înregistrate în tabela de numere întîmplătoare.

Doi statisticieni americani, Metropolis şi Ulam, folosind numerele întîmplătoare rezultate de la ruleta din cazinoul Monte Carlo, au dat acestor procedee de calcul numele de „metoda Monte Carlo". Acest nume s-a răspîndit foarte mult în literatura de specialitate. Noi renunţăm la el pentru motivul că numerele date de ruleta de la Monte Carlo reprezintă un caz particular de numere întîmplătoare, care astăzi nici nu mai sînt folosite în ştiinţă. Cu ajutorul maşinilor electronice de calcul, s-au întocmit noi tabele de numere întîmplătoare, mai corespunzătoare din punct de vedere ştiinţific, care sînt folosite pe scară largă.

2. Calculul valorii lui π

Să considerăm schema lui Bernoulli constînd din extracţii repetate de bile dintr-o urnă de compoziţie cunoscută, punînd, de fiecare dată, bila extrasă înapoi în urnă. Dacă în urnă sînt bile albe şi bile negre, să notăm prin p probabilitatea scoaterii unei bile albe şi prin q probabilitatea scoaterii unei bile negre ($p + q = 1$). Fie α numărul de bile albe obţinut în n extracţii consecutive şi $l = \alpha - n \cdot p$, abaterea corespunzătoare. Am arătat că

$$M(l^2) = npq. \qquad \text{(XXIII, 2.1)}$$

Se poate arăta că

$$M(|l|) = \sqrt{2\pi npq}. \qquad \text{(XXIII, 2.2)}$$

O verificare rapidă a acestei formule se poate obţine din expresia asimptotică a probabilităţii de a obţine o abatere l. Într-adevăr, am arătat că

$$p_l \cong \frac{1}{\sqrt{2\pi npq}} \, e^{-\frac{l^2}{2npq}}. \qquad \text{(XXIII, 2.3)}$$

În mod aproximativ avem

$$M(|l|) = \frac{1}{\sqrt{2\pi npq}} \int_{-\infty}^{+\infty} |l| \, e^{-\frac{l^2}{2npq}} \, dl =$$

$$= \frac{2}{\sqrt{2\pi npq}} \int_{0}^{+\infty} l e^{-\frac{l^2}{2npq}} \, dl = \sqrt{2\pi npq}. \qquad \text{(XXIII, 2.4)}$$

Pe de altă parte,

$$M[|l|^2] = M(l^2) = npq, \qquad \text{(XXIII, 2.5)}$$

conform formulei (XXIII, 2.1).

Să considerăm acum s serii de extracții independente făcute din urna de mai sus. Fiecare serie conține n extracții. Notăm prin

$$l_1, l_2, \ldots, l_s$$

abaterile obținute în aceste serii.

Abaterile absolute:

$$|l_1|, |l_2|, \ldots, |l_s|$$

pot fi considerate ca o selecție de n probe independente făcute dintr-o populație teoretică a cărei medie și dispersie sînt date respectiv de formulele (XXIII, 2.2) și (XXIII, 2.1).

Să considerăm expresia

$$L_s = \frac{|l_1| + |l_2| + \ldots + |l_s|}{s}. \qquad \text{(XXIII, 2.6)}$$

Se știe din teoria selecției că L_s pentru $s \to \infty$, tinde către o lege normală, bineînțeles după ce L_s a fost supus unei normări convenabile. Media lui L_s este egală cu m, iar dispersia lui cu $\frac{\sigma}{\sqrt{s}}$; prin m și σ însemnînd respectiv media (XXIII, 2.2) și dispersia (XXIII, 2.1). Drept consecință cu o probabilitate mai mare decît 0,997 avem îndeplinite inegalitățile:

$$\sqrt{2\pi npq} - \frac{3}{\sqrt{s}}\sqrt{npq} \leqslant L_s \leqslant \sqrt{2\pi npq} + \frac{3}{\sqrt{s}}\sqrt{npq} \qquad \text{(XXIII, 2.7)}$$

sau

$$-\frac{3}{\sqrt{2s}} \leqslant \frac{L_s}{\sqrt{2npq}} - \sqrt{\pi} \leqslant \frac{3}{\sqrt{2s}}. \qquad \text{(XXIII, 2.8)}$$

$\frac{L_s}{\sqrt{2npq}}$ este o valoare experimentală și se apropie de $\sqrt{\pi}$ în baza formulei de mai sus, cu cît numărul experiențelor s este mai mare. Se vede că diferența dintre valoarea experimentală și $\sqrt{\pi}$ este de ordinul lui $\frac{1}{\sqrt{s}}$. De exemplu dacă am face 100 000 de experiențe găsim valoarea lui $\sqrt{\pi}$ cu primele două cifre exacte. Pentru a calcula pe $\sqrt{\pi}$ cu un număr mare de zecimale exacte trebuie deci să facem un număr foarte mare de experiențe, ceea ce cu ajutorul tabelelor de numere întîmplătoare se poate ușor realiza.

Czuber*) a realizat experiența de mai sus pentru $s = 37$, $n = 100$, și $p = \frac{1}{2}$. În cele 37 de serii de experiențe pe care le-a făcut a găsit următoarele abateri:

$-8, -7, -6, -5, -4, -3, -2, 1, 0, +1, +2, +3, +4, +5, +6, +7, +8,$

cu frecvențele următoare:

$$1, 1, 2, 2, 2, 4, 1, 4, 4, 2, 4, 1, 2, 0, 5, 0, 2.$$

*) E. Czuber, *Wahrscheinlichkeitsrechnung*, vol. I, p. 330.

Pentru L_s rezultă valoarea $L_{37} = 3{,}51$. Se vede ușor că intervalul de încredere (XXIII, 2.7) este verificat. Valoarea lui $\sqrt{\pi}$ astfel obținută este cu 1,25% mai mare decît în realitate din cauză că numărul seriilor, $s = 37$, este foarte mic.

3. Problema acului

Buffon (1877) a enunțat și rezolvat celebra problemă a acului din calculul probabilităților.

Pe o placă orizontală sînt desenate un șir de drepte paralele, egal depărtate. Se aruncă peste această placă un ac a cărui lungime este mai mică decît distanța dintre drepte. Care este probabilitatea ca acul să întîlnească una dintre drepte?

În această problemă vom avea de evaluat măsura mulțimii de poziții diferite pe care le poate lua acul și din aceasta a evalua deosebit măsura mulțimii de poziții în care acul întîlnește una din drepte. Probabilitatea căutată va fi raportul acestor două măsuri. Cum vom avea grijă ca măsura cea dintîi să fie 1, rezultatul va fi dat chiar de măsura a doua.

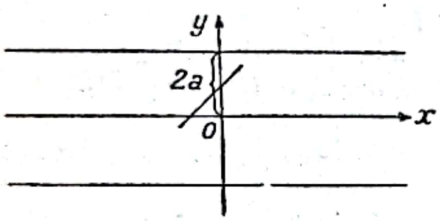


Fig. 49

Să alegem un punct O pe una din drepte și prin acest punct să ducem o perpendiculară comună. Vom presupune că mijlocul acului se găsește pe această dreaptă. Potrivit unui principiu pe care l-am mai întrebuințat, restricția aceasta nu va schimba valoarea probabilității căutate. De asemenea, în baza aceluiași principiu vom restrînge probabilitatea de variație a ordonatei mijlocului, la intervalul O, a și atunci probabilitatea ca acest punct să se găsească între y și $y + dy$ este

$$\frac{dy}{a}.$$

Probabilitatea ca acul să facă cu Oy un unghi cuprins între α și $\alpha + d\alpha$ este

$$\frac{d\alpha}{\pi}.$$

Deci, probabilitatea ca acul să aibă mijlocul între y și $y + dy$ și orientarea între α și $\alpha + d\alpha$ este

$$\frac{d\alpha \cdot dy}{a\pi}.$$

Măsura totalității pozițiilor este

$$\frac{1}{a\pi}\int_0^a \int_{-\frac{\pi}{2}}^{+\frac{\pi}{2}} d\alpha \cdot dy = 1.$$

Măsura mulțimii de poziții pentru care acul întîlnește dreapta Ox este

$$\frac{1}{a\pi}\iint_D da \cdot dy,$$

domeniul la care se extinde integrala fiind totalitatea pozițiilor pentru care $y \leqslant l \cos a$, $2l$ fiind lungimea acului.

Aria exprimată prin integrala dublă de mai sus este limitată de axa Oa și de curba $y = l \cos a$, cînd a variază de la $-\frac{\pi}{2}$ la $\frac{\pi}{2}$. Această arie este

$$\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} l \cos a \cdot da = 2l.$$

Deci probabilitatea căutată este

$$P = \frac{2l}{a\pi}.$$

Calculul precedent ne dă mijlocul de a evalua imediat lungimea unei curbe convexe.

Pentru a trece la atari considerații să observăm mai întîi că rezultatul precedent poate fi interpretat astfel:

Măsura ansamblului de segmente de lungime $2l$ care întîlnește una din dreptele rețelei, este proporțională cu lungimea.

Să considerăm o linie curbă (în aceleași condiții ca acul, adică astfel ca să nu poată întîlni două linii în același timp).

Să privim curba ca o linie poligonală. Suma măsurilor respective, pentru fiecare dintre laturile corespunzătoare, adică ($2\lambda_i$ fiind una din laturi) este:

$$\frac{\Sigma 2\lambda_i}{a\pi} = \frac{L}{a\pi},$$

unde L este lungimea curbei.

Măsura mulțimii de poziții ale curbei în care curba întîlnește o singură dată una din drepte plus de două ori măsura mulțimii de poziții pentru care curba întîlnește una din drepte în două puncte plus de trei ori măsura mulțimii de poziții pentru care curba întîlnește una din drepte în trei puncte și așa mai departe, este:

$$1.\ P_1 + 2P_2 + \ldots + kP_k + \ldots = \frac{L}{a\pi}.$$

Considerațiile acestea se verifică, de exemplu, în cazul unui cerc de diametru $2a$. Un atare cerc întîlnește cu siguranță una din drepte

și anume în două puncte. Vom avea deci $P_2 = 1$. Formula precedentă dă

$$2 = \frac{L}{a\pi}$$

de unde

$$L = 2a\pi,$$

formula lungimii cercului.

Să luăm acum exemplul mai general al unei curbe convexe închise oarecare. Fie P_2 probabilitatea ca o atare curbă să întîlnească una din dreptele rețelei. Vom avea

$$P_2 = \frac{L}{2a\pi}$$

sau încă

$$L = 2a\pi \cdot P_2.$$

Această probabilitate P_2 se poate stabili în modul următor:

Dacă considerăm un cerc de rază a care cuprinde în interiorul său curba și ducem o coardă a cercului, P_2 poate fi interpretat ca probabilitatea ca această coardă a cercului să întîlnească curba și este egală cu raportul lungimii curbei la aceea a cercului.

Dacă notăm cu p distanța coardei de centrul cercului și cu a unghiul pe care perpendiculara dusă la coardă din O îl face cu o direcție dată, probabilitatea elementară este

$$\frac{\mathrm{d}p \cdot \mathrm{d}a}{2a\pi}.$$

Probabilitatea ca o coardă să întîlnească curba C este dată de relația

$$\frac{1}{2a\pi}\int_0^{2\pi} \mathrm{d}a \cdot \int_0^{r(a)} \mathrm{d}p = \frac{1}{2a\pi}\int_0^{2\pi} r(a) \cdot \mathrm{d}a,$$

unde $r(a)$ este distanța de la O la tangenta corespunzătoare direcției a; vom avea deci

$$P_2 = \frac{L}{2a\pi} = \frac{1}{2a\pi}\int_0^{2\pi} r(a)\ \mathrm{d}a,$$

de unde

$$L = \int_0^{2\pi} r(a) \cdot \mathrm{d}a,$$

formulă găsită pe altă cale de Cauchy.

Scanned with OKEN Scanner

În baza teoremei lui Bernoulli, pentru ca, cu o probabilitate egală cu 0,90, expresia

$$\left| \frac{\alpha}{n} - \frac{2}{\pi} \right|,$$

(unde α reprezintă numărul de cîte ori acul atinge o linie în n probe) să fie mai mică decît 0,01, ar trebui să se facă 6000 de experienţe. Aceasta ar conduce la o evaluare a lui π cu o eroare mai mică decît 0,05.

Experienţele făcute au confirmat teoria. Astronomul Wolf, cu ajutorul a 5 000 de probe a găsit pentru π valoarea 3,1596, M. A. Smith (1855) luînd $n = 3\,204$ a găsit valoarea 3,1553, Lazzerini (1902) făcînd 2 000 experienţe a obţinut $\pi = 3{,}1446$.

Problema acului a fost rezolvată cu ajutorul unor distribuţii continue, care, în calculul probabilităţilor au fost denumite probabilităţi geometrice.

Pentru împrejurările cînd evenimentele, cărora le atribuim probabilităţi, formează o mulţime infinită, este nevoie de a lărgi definiţia probabilităţii, fără a-i altera nimic din ceea ce este caracteristic în cazul unui număr finit de cazuri. Acest procedeu a fost aplicat pentru rezolvarea problemei acului a lui Buffon. Numele de probabilităţi geometrice vine de la faptul că fiecare eveniment poate fi asimilat cu un punct al unei linii, al unei suprafeţe, al unui volum etc. şi că toate mulţimile de evenimente sînt entităţi geometrice. J. Bertrand a atras cel dintîi atenţia asupra dificultăţilor ce se pot întîlni la definirea exactă a probabilităţilor geometrice. El şi-a propus problema să se determine probabilitatea coardelor unui cerc de a fi mai mari ca latura triunghiului echilateral înscris. După modul de definire al probabilităţii geometrice respective, Bertrand a găsit pentru probabilitatea căutată mai multe valori şi anume: 1/3, 1/2, $1 - \frac{\sqrt{3}}{2}$ etc. Se ajunge la faptul absurd că pentru acelaşi eveniment să existe mai multe probabilităţi. Exemplul de mai sus este cunoscut în istoria matematicii sub numele de *paradoxul lui Bertrand*. El a făcut pe mulţi matematicieni de la sfîrşitul secolului al XIX-lea să se îndoiască de exactitatea teoriei probabilităţilor. Considerarea probabilităţii ca o măsură a unei mulţimi de evenimente, a înlăturat obiecţiile din paradoxul lui Bertrand.

Începînd din anul 1935, W. Blaschke a publicat o serie de studii de probabilităţi geometrice în care predomină însă caracterul geometric, fapt pentru care el le-a numit „geometrie integrală". Pe linia trasată de Blaschke au lucrat ulterior numeroşi matematicieni printre care citäm pe cei care au adus cele mai importante contribuţii şi anume: L. A. Santalo şi S. S. Chem, care au extins geometria integrală la spaţii mai generale, şi pe matematicienii sovietici I. M. Iaglom şi pe P. K. Raschevski, care au folosit ecuaţiile diferenţiale ale unei familii de curbe din plan. La noi în ţară, M. Stoka a obţinut o serie de rezultate în domeniul geometriei integrale. Expunerea lucrărilor lui Stoka iese din cadrul de specialitate al acestei cărţi. Ne mulţumim numai să arătăm că în cazul cînd evenimentele sînt elemente geometrice, destul de generale, el a reuşit să găsească probabilităţile geometrice corespunzătoare ca de exemplu:
pentru mulţimea cercurilor din plan

$$(x - a)^2 + (y - b)^2 = c,$$

pentru mulţimea conicelor din plan:

$$a_{11}x^2 + 2a_{12}xy + a_{22}y^2 + 2a_{13}x + 2a_{23}y + 1 = 0$$

şi aşa mai departe.

4. Folosirea numerelor întîmplătoare

Exemplele de la punctul 1 și 2 sînt clasice în teoria probabilităților. Ele dovedesc că ideea folosirii convergenței statistice pentru calculul unor valori este destul de veche. După cum am mai spus, contribuția esențială pe care a adus-o știința actuală în realizarea practică a acestei idei este folosirea „numerelor întîmplătoare".

Există numeroase tabele de numere întîmplătoare obținute pe diferite căi. La început s-au întocmit tabele din rezultatele loteriilor, din amestecul jocurilor de cărți etc. S-a renunțat însă în curînd la asemenea procedee, care nu ne pot procura cu ușurință seriile lungi de numere întîmplătoare de care avem nevoie în aplicațiile practice. S-a recurs de aceea la diferite metode din fizică, de exemplu la întrebuințarea unor aparate speciale cu care se calculează numărul de vibrații corespunzător unor impulsuri, sau la mașinile electronice de calcul. Să arătăm un procedeu cu ajutorul căruia putem calcula numere întîmplătoare cu patru cifre. Pornim de la un număr arbitrar de patru cifre

$$X_0 = 3987;$$

prin ridicare la pătrat rezultă

$$X_0^2 = 15896169.$$

Ne alegem al doilea număr întîmplător X_1, numărul de patru cifre format din considerarea cifrelor mijlocii ale lui X_0^2 adică

$$X_1 = 8961.$$

Repetînd același procedeu avem

$$X_1^2 = 80299521.$$

Deci al treilea număr întîmplător este

$$X_2 = 2995.$$

Pentru determinarea numărului întîmplător care urmează, ridicăm din nou la pătrat și obținem

$$X_2^2 = 08970025.$$

Rezultă astfel numărul întîmplător

$$X_3 = 9700.$$

Am obținut pînă acum numerele întîmplătoare cu patru cifre: 3987, 8961, 2995 și 9700. Pe aceeași cale putem să completăm tabelul de numere întîmplătoare.

De multe ori probleme de natură specială cer ca numerele întîmplătoare să fie scrise într-o bază de numerotație diferită de 10. Numerele

întîmplătoare de acest gen se obțin imediat scriind numerele întîmplătoare din baza 10 în baza corespunzătoare. De exemplu:

$$3987 = 1.2^{11} + 1.2^{10} + 1.2^{9} + 1.2^{8} + 1.2^{7} + 0.2^{6} + 0.2^{5} + 1.2^{4} + 0.2^{3} + \\ + 0.2^{2} + 1.2^{1} + 0.1,$$

$$8961 = 1.2^{13} + 0.2^{12} + 0.2^{11} + 0.2^{10} + 1.2^{9} + 1.2^{8} + 0.2^{7} + 0.2^{6} + 0.2^{5} + \\ + 0.2^{4} + 0.2^{3} + 0.2^{2} + 0.2 + 0.1,$$

$$2995 = 1.2^{11} + 0.2^{10} + 1.2^{9} + 1.2^{8} + 1.2^{7} + 0.2^{6} + 1.2^{5} + 1.2^{4} + 0.2^{3} + \\ + 0.2^{2} + 1.2 + 0.1,$$

$$9700 = 1.2^{13} + 0.2^{12} + 0.2^{11} + 1.2^{10} + 0.2^{9} + 1.2^{8} + 1.2^{7} + 1.2^{6} + 1.2^{5} + \\ + 0.2^{4} + 0.2^{3} + 1.2^{2} + 0.2 + 0.1.$$

Prin urmare, numerele întîmplătoare 3987, 8961, 2995 și 9700 din baza zece devin următoarele numere întîmplătoare în baza 2:

111 110 010 011;

10 001 100 000 001;

101 110 110 011;

10 010 111 100 100.

Ultimele numere întîmplătoare ne pot servi la determinarea seriilor de frecvențe de la exemplul tratat la punctul 1 în cazul cînd probabilitatea de a extrage o bilă albă este 0,5. Dacă avem nevoie de serii mai mari va fi suficient să considerăm numere întîmplătoare cu mai multe cifre. Astfel numerele întîmplătoare cu șase cifre, în baza 10, ne conduc la serii de numere întîmplătoare cu 20 de cifre, scrise în baza 2.

În cazul distribuțiilor discrete, există însă un procedeu general pentru a obține serii de frecvențe datorită întîmplării. Astfel, în cazul examinat la punctul 1, ne propunem să obținem o serie de bile albe și negre ieșite la întîmplare în 20 de extracții făcute din urnă. Admitem că probabilitatea de a scoate o bilă albă este $p = 0{,}35$ și drept consecință probabilitatea de a scoate o bilă neagră este $q = 0{,}65$. Folosim un tabel de numere întîmplătoare cu două cifre. Esențial în construirea numerelor întîmplătoare este faptul că toate aceste numere au aceeași probabilitate de a se ivi. Numerele întîmplătoare corespund unei distribuții uniforme. Putem deci admite că probabilitatea de a avea un număr întîmplător mai mic decît 35 este egală cu 0,35 și probabilitatea de a obține un număr întîmplător mai mare sau egal cu 35 este 0,65.

Dacă obținem deci un număr întîmplător mai mic decît 35, să zicem 19, înseamnă că s-a realizat un eveniment de probabilitate 0,35 adică echivalentul scoaterii unei bile albe din urna noastră.

Dacă notăm prin 1 extracția unei bile albe și prin 0 extracția unei bile negre putem întocmi pe baza tabelului de numere întîmplătoare de două cifre, următoarea schemă:

Nr. experienței	Nr. întîmplător	Rezultatul experienței
1	83	0
2	36	0
3	63	0
4	85	0
5	91	0
6	87	0
7	93	0
8	94	0
9	10	1
10	42	0
11	14	1
12	03	1
13	35	0
14	75	0
15	42	0
16	88	0
17	13	1
18	23	1
19	41	0
20	23	1

În această experiență de 20 de extracții ne-au ieșit 6 bile albe și 14 bile negre. Experiența poate fi continuată cu un număr oricît de mare de numere întîmplătoare, obținînd astfel serii de frecvență de orice lungime. Se vede inutilitatea continuării executării experienței fizice constînd din scoaterea unei bile din urnă. Numerele întîmplătoare obținute din experiențele trecute ne conduc la frecvențe care, din punct de vedere probabilistic, sînt echivalente cu rezultatele experiențelor actuale.

O lipsă de precizie în alcătuirea unui astfel de tablou se întîlnește atunci cînd numărul întîmplător este foarte aproape de probabilitatea p. Pentru a vedea cum se înlătură această dificultate vom reface tabloul de mai sus în ipoteza că p este egal cu 0,358. Numerele întîmplătoare putem să le scriem și sub formă zecimală. De exemplu numărul întîmplător 85 poate fi scris — sub formă echivalentă — 0,85. Toate numerele întîmplătoare mai mici decît 0,358, înseamnă realizarea evenimentului de probabilitate $p = 0{,}358$. Cum numerele întîmplătoare sînt cu două zecimale, înseamnă că toate numerele întîmplătoare de la 0 —0,34 inclusiv, corespund realizării evenimentului. Rămîne un caz de ambiguitate atunci cînd numărul întîmplător este egal cu 0,35, cum este cazul în rîndul 13 din tabloul de mai sus. În acest caz, la sfîrșitul numărului zecimal 0,35 adăugăm cifrele numărului întîmplător următor, care după cum se vede din rîndul 14, este egal cu 75. Obținem astfel numărul întîmplător 0,3575, care este mai

mic decît $p = 0{,}358$. Înseamnă, prin urmare, că evenimentul s-a realizat și deci în linia 13 și coloana a treia trebuie de data aceasta să fie cifra 1 în loc de 0. Restul tabloului rămîne neschimbat. Prin urmare pentru $p = 0{,}35$ dintr-o serie de 20 de experiențe evenimentul s-a realizat de 6 ori și de 14 ori nu s-a realizat, pe cînd pentru $p = 0{,}358$ în același număr de experiențe s-a realizat evenimentul de 7 ori și de 13 ori nu s-a realizat.

Este important să cunoaștem modul de construire a numerelor întîmplătoare pentru o lege de probabilitate continuă.

Fie $f(x)$ densitatea de probabilitate a unei distribuții continue:

$$f(x) \geqslant 0, \int_{-\infty}^{+\infty} f(x)\mathrm{d}x = 1.$$

Ne propunem că calculăm numere întîmplătoare corespunzătoare acestei distribuții pornind de la un tabel obișnuit de numere întîmplătoare. Considerăm relația

$$Y = \int_{-\infty}^{X} f(x)\mathrm{d}x \qquad \text{(XXIII, 4.1)}$$

pe care o punem și sub forma

$$Y = H(X), \qquad \text{(XXIII, 4.2)}$$

$H(X)$ reprezintă prin urmare operația (XXIII, 4.1). Admitem că Y, X, legați prin operația (XXIII, 4.1) sau (XXIII, 4.2), sînt variabile aleatoare. Distribuția lui Y este o distribuție uniformă, adică dacă notăm prin $F(y)$ funcția de repartiție a variabilei Y avem:

$$F(y) = P[Y \leqslant y] = y. \qquad \text{(XXIII, 4.3)}$$

Fie $G(x)$ funcția de repartiție a variabilei X:

$$G(x) = P(X \leqslant x) = P(H^{-1}(Y) \leqslant x) = P[Y \leqslant H(x)] = H(x). \qquad \text{(XXIII, 4.4)}$$

Egalitățile de mai sus, unde $H^{-1}(x)$ este funcția inversă a lui $H(x)$, pot fi scrise, deoarece $F(x)$, $G(x)$, $H(x)$ sînt funcții continue, monoton crescătoare.

Din (XXIII, 4.4) și (XXIII, 4.1) deducem atunci

$$P(X \leqslant x) = \int_{-\infty}^{X} f(x)\mathrm{d}x. \qquad \text{(XXIII, 4.5)}$$

Am stabilit astfel următoarea proprietate interesantă: dacă Y este o variabilă aleatoare uniformă, atunci X este o variabilă aleatoare corespunzătoare densității de probabilitate $f(x)$. În baza relației (XXIII, 4.1) se mai vede că Y ia numai valori pozitive cuprinse între 0 și 1.

În baza acestei proprietăți, se pot construi ușor numere întîmplătoare corespunzătoare densității $f(x)$ plecînd de la un tabel obișnuit de numere întîmplătoare. Ținînd seama că Y este un număr zecimal, cuprins între 0 și 1, toate numerele întîmplătoare vom presupune că reprezintă partea zecimală a unui număr pozitiv mai mic sau egal cu 1. De

exemplu: numărul întîmplător 89 ne conduce la valoarea $Y = 0,89$ şi pentru determinarea lui X avem de rezolvat ecuaţia

$$0,89 = \int_{-\infty}^{X} f(x)\mathrm{d}x.$$

Să calculăm prin acest procedeu un şir de numere întîmplătoare corespunzătoare legii normale $N(0,1)$, adică — după cum ştim — legii normale care are media 0 şi dispersia 1. În acest caz avem

$$f(x) = \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-\frac{x^2}{2}}.$$

Valorile funcţiei

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} \mathrm{e}^{-\frac{x^2}{2}} \mathrm{d}x \qquad \text{(XXIII, 4.6)}$$

sînt cunoscute din tabele special construite pentru legea normală.

Să construim 20 de numere întîmplătoare corespunzătoare legii normale $N(0,1)$ pornind de la un tabel de numere întîmplătoare cu două cifre.

Formăm tabelul 117.

Tabelul 117

Numărul experienţei	Numărul întîmplător	Frecvenţă corespunzătoare legii normale
1	2	3
1	0,83	0,96
2	0,36	−0,35
3	0,63	0,34
4	0,85	1,04
5	0,91	1,34
6	0,87	1,13
7	0,93	1,48
8	0,94	1,56
9	0,10	−1,28
10	0,42	−2,00
11	0,14	−1,08
12	0,03	−1,88
13	0,35	−0,38
14	0,75	−0,68
15	0,42	−2,00
16	0,88	1,17
17	0,13	−1,12
18	0,23	−0,74
19	0,41	−0,23
20	0,23	−0,74

Ultima coloană ne dă frecvenţele corespunzătoare legii normale $N(0,1)$. Avem 8 frecvenţe pozitive şi 12 frecvenţe negative. În intervalul cuprins între prima cuartilă teoretică şi a treia cuartilă teoretică, adică în intervalul $(-0,68 + 0,68)$, avem 11 frecvenţe. Nu avem nici o abatere

care în valoare absolută să fie mai mare de trei. După cum se vede, primele verificări aplicate la frecvențele legii normale $N(0,1)$ dau rezultate satisfăcătoare. Numerele întîmplătoare ale unei distribuții normale $N(m, \sigma)$ se obțin din numerele de mai sus cu ajutorul relației

$$x' = \frac{x - m}{\sigma}.$$

5. Calculul unei arii mărginite de o curbă convexă

Notăm cu A aria mărginită de curba convexă avînd drept ecuație

$$a(x, y) = 0.$$

Fără ca să restrîngem generalitatea, putem presupune că această curbă este cuprinsă în pătratul ale cărui vîrfuri au coordonate (0,0), (1,0), (0,1), (1,1) (fig. 50).

Considerăm un proces aleator constînd dintr-o ploaie uniformă de particule care cad pe pătrat. Probabilitatea elementară a particulelor este

$$dP = dx\, dy$$

și probabilitatea ca o particulă să cadă în aria A este

$$P = \iint_{(A)} dx\, dy.$$

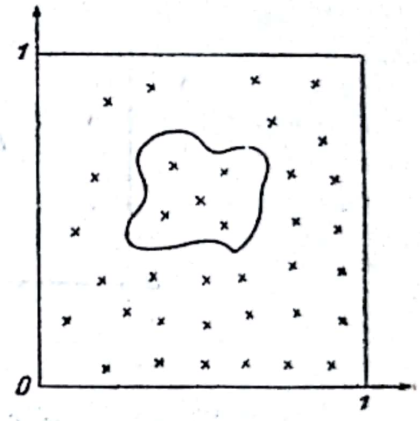


Fig. 50

Evident, probabilitatea ca o particulă să cadă în pătrat este egală cu 1. Vom aplica metoda statistică de calcul, înlocuind experiența cu o lectură de numere întîmplătoare cuprinse între 0 și 1. Două numere consecutive le considerăm drept coordonatele unui punct în plan. Urmărind pe tabel, vom număra cîte perechi de numere dau puncte ce cad în interiorul curbei convexe considerate. Să presupunem că din N puncte au căzut în interiorul curbei n puncte. Numărul $\frac{n}{N}$ ne dă o valoare aproximativă a ariei A, care este cu atît mai exactă cu cît numărul numerelor întîmplătoare folosite este mai mare.

6. Calculul unei integrale definite

Ne propunem să calculăm integrala

$$I = \int_a^b f(x)dx,$$

unde $f dx$ este o funcție cunoscută.

Reprezentăm grafic funcția $y = f(x)$ (fig. 51).

Valoarea integralei este egală cu aria haşurată. Ajungem prin urmare la problema tratată la punctul precedent care se rezolvă prin considerarea unui proces aleator constînd dintr-o ploaie uniformă de particule.

Ne propunem să calculăm o integrală de forma

$$\int_{-\infty}^{+\infty} g(x)f(x)\mathrm{d}x, \qquad \text{(XXIII, 6.1)}$$

unde $f(x)$ este o densitate de probabilitate, adică

$$f(x) \geqslant 0, \quad \int_{-\infty}^{+\infty} f(x)\mathrm{d}x = 1.$$

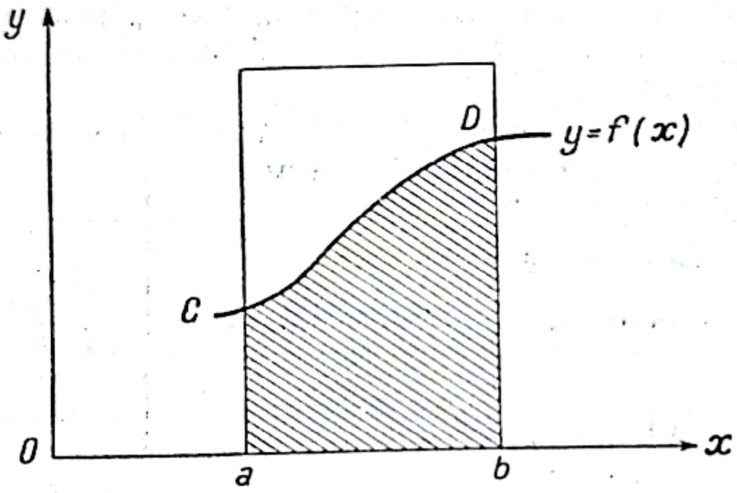


Fig. 51

Presupunem că avem tabele de numere întîmplătoare corespunzătoare densității de probabilitate $f(x)$ şi că în baza lor executăm o selecție de volum N:

$$X_1, X_2, \ldots, X_n.$$

Formula (XXIII, 6.1) ne arată că integrala a cărei valoare nu ne este cunoscută, este valoarea medie a funcției $g(x)$. Din teoria selecției ştim că media de selecție

$$\frac{1}{N}\sum_{i=1}^{N} g(x_i) \qquad \text{(XXIII, 6.2)}$$

este o funcție de estimație absolut corectă a valorii mediei teoretice. Prin urmare, pentru valorile mari ale lui N, expresia (XXIII, 6.2) este foarte apropiată de (XXIII, 6.1).

Pentru determinarea erorii de aproximare se utilizează dispersia. Variabilele $X_i, (i = 1, 2, \ldots, N)$, sînt independente între ele. Prin urmare

$$D^2\left(\frac{1}{N}\sum_{i=1}^{N} X_i\right) = \frac{\sigma_g^2}{N}, \qquad \text{(XXIII, 6.3)}$$

unde prin σ_g^2 am notat dispersia teoretică a funcției $g(x)$.

Această egalitate ne arată că dispersia funcției de selecție (XXIII, 6.2) este de ordinul lui $N^{-\frac{1}{2}}$. Ca un caz particular să considerăm o distribuție uniformă de probabilitate și integrala

$$I = \int_0^1 g(x)\mathrm{d}x; \qquad \text{(XXIII, 6.4)}$$

în baza metodei mai sus arătate, valoarea integralei I este dată cu aproximație de media de selecție

$$\frac{1}{N}\sum_{i=1}^{N} g(x_i). \qquad \text{(XXIII, 6.5)}$$

Să comparăm formula de mai sus cu formula clasică a integrării numerice în baza căreia I este dat cu aproximație de media aritmetică a valorilor $g(x)$, calculate în punctele $\frac{1}{2N}, \frac{3}{2N}, \frac{5}{2N}, \ldots$ Prin această metodă intervalul (0,1) se împarte în N părți egale și se consideră valorile funcției $g\mathrm{d}x$ în mijlocul fiecărui subinterval. Deosebirea esențială între metoda clasică și metoda statistică constă în faptul că alegerea punctelor în formula (XXIII, 6.5) se face sistematic în prima metodă și la întîmplare în metoda a doua. Cu alte cuvinte, în calculul clasic al integralei, bazat pe definiția lui Riemann, folosim o alegere sistematică de puncte, pe cînd prin metoda statistică folosim o alegere întmplătoare. În general alegerea întîmplătoare dă naștere la erori mult mai mici, deoarece erorile făcute la fiecare punct prin însumare se pot compensa.

Dar superioritatea alegerii întîmplătoare apare la calculul integralelor multiple. Ca exemplu să considerăm funcția continuă $g(x_1, x_2, \ldots, x_k)$ și să calculăm integrala

$$\int\int\ldots\int_{(C)} g(x_1, x_2, \ldots, x_k)\,\mathrm{d}x_1\mathrm{d}x_2 \ldots \mathrm{d}x_k.$$

prin C însemnînd un cub cu k dimensiuni.

După metoda clasică este necesar să construim o rețea cu dimensiunea celulelor $\frac{1}{m}$, și să calculăm valoarea funcției $g(x_1, x_2, \ldots, x_k)$ în m^k punte.

Dacă luăm m egal numai cu 30, iar $k = 6$, atunci $m^k = 729\,000\,000$. Se demonstrează că dacă celulele alese sînt cuburi egale și plasăm nodurile de integrale în centrul acestor cuburi, eroarea obținută este de ordinul lui m^{-2}. Dacă am fi folosit o alegere întîmplătoare ca să obținem aceeași eroare, ar fi trebuit să folosim o selecție de mărime N astfel ca

$$N^{-\frac{1}{2}} = m^{-2}$$

sau

$$N = m^4.$$

În cazul nostru, m fiind egal cu 30, obținem $N = 810\,000$. Prin urmare pentru a obține valoarea integralei cu aceeași eroare de aproximație, trebuie să calculăm prin metoda clasică valoarea funcției în 729 000 000 de puncte, iar prin metoda statistică în 810 000 de puncte, ceea ce constituie o diferență apreciabilă chiar și pentru mașinile electronice de calculat. Superioritatea folosirii metodei statistice este manifestă.

7. Mersul la întîmplare

Împărțim un interval $(0, a)$ într-un număr de părți egale cu lungimea h. Pe segment se află o particulă care se deplasează cu distanța h fie la dreapta, fie la stînga, pe baza unui proces aleator. Dacă particula se află la un moment dat în punctul x, probabilitatea ca să meargă la dreapta, adică în punctul $x + h$, este p și probabilitatea ca să meargă la stînga, în punctul $x - h$, este q, $(p + q = 1)$.

Procesul aleator de mișcare la întîmplare se termină cînd particula ajunge într-unul din punctele 0 sau a.

Să notăm prin $V(x)$ probabilitatea ca particula pornind din punctul x să ajungă în a. Este evident că avem $V(0) = 0$ și $V(a) = 1$. Funcția $V(x)$ verifică ecuația cu diferențe finite

$$V(x) = pV(x + h) + qV(x - h) \qquad \text{(XXIII, 7.1)}$$

cu condițiile la limită

$$V(0) = 0, \qquad V(a) = 1.$$

Relația (XXIII, 7.1) este ușor de dovedit pe baza proprietății de adunarea probabilităților. Într-adevăr, particula poate ajunge în a în două moduri incompatibile între ele. Ea poate să facă un pas la dreapta și apoi din punctul $x + h$ să ajungă în a, ceea ce se întîmplă cu probabilitatea

$$pV(x + h) \qquad \text{(XXIII, 7.2)}$$

sau poate să facă un pas la stînga și apoi din punctul $x - h$ să ajungă în a ceea ce se întîmplă cu probabilitatea

$$qV(x - h). \qquad \text{(XXIII, 7.3)}$$

Adunînd probabilitățile (XXIII, 7.2) și (XXIII, 7.3), obținem probabilitatea evenimentului total care este $V(x)$.

Vom pune relația (XXIII, 7.1) sub o formă mai convenabilă. Mai întîi din faptul că $p + q = 1$, rezultă

$$(p + q)V(x) = qV(x - h) + pV(x + h)$$

sau

$$V(x + h) + V(x - h) - 2V(x) - \frac{q - p}{q}[V(x + h) - V(x)] = 0.$$

Împărțim ambii membri ai acestei relații cu h^2 și notăm

$$\alpha = \frac{q-p}{qh}.$$

Obținem astfel

$$\frac{V(x+h)+V(x-h)-2V(x)}{h^2} - \alpha\frac{V(x+h)-V(x)}{h} = 0. \qquad \text{(XXIII, 7.4)}$$

Dacă p este o funcție de h, astfel ca

$$\frac{q-p}{q} = \alpha h + 0(h),$$

unde

$$\frac{0(h)}{h} \to 0, \quad \text{pentru } h \to 0,$$

atunci pentru $h \to 0$, ecuația devine

$$\frac{d^2V}{dx^2} - \alpha\frac{dV}{dx} = 0, \qquad \text{(XXIII, 7.5)}$$

adică o ecuație diferențială de ordinul al doilea, cu condițiile la limită

$$V(0) = 0, \quad V(a) = 1.$$

Ajungem deci la posibilitatea integrării ecuației diferențiale (XXIII, 7.5) prin metode statistice.

Cu ajutorul mașinilor electronice de calcul, pe baza unui tabel de numere întîmplătoare, se execută un număr N de mersuri la întîmplare, pornind din același punct x. Dacă însemnăm cu n numărul de opriri în punctul a, avem $V(x) \approx \frac{N}{n}$ și obținem soluția ecuației în punctul x. Pentru a înțelege mai bine modul de aplicare al acestui procedeu, vom expune un exemplu concret de rezolvare a unei ecuații diferențiale de tipul (XXIII, 7.5) tratat de Antoneta Popescu în lucrarea sa de diplomă pentru examenul de diplomă la Facultatea de matematică și fizică din București, sesiunea iunie 1960. Se pornește de la ecuația diferențială:

$$y'' - \frac{40}{11}y' = 0 \qquad \text{(XXIII, 7.6)}$$

cu condițiile la limită

$$y(0) = 0, \quad y(1) = 1. \qquad \text{(XXIII, 7.7)}$$

Se propune să se calculeze valoarea soluției acestei ecuații diferențiale în punctul $x = 0{,}8$. Soluția acestei ecuații diferențiale este

$$y = \frac{1}{36{,}985}(e^{\frac{40}{11}x} - 1).$$

Pentru $x = 0{,}8$, această formulă dă

$$y(0{,}8) = 0{,}404.$$

Deci calculul direct dă valoarea 0,404 pentru soluția căutată.

Să procedăm acum prin metoda statistică. Considerăm intervalul [0,1] împărțit în 20 de părți egale. Admitem prin urmare că pasul $h = \frac{1}{20}$.

Se presupune că particula pleacă din punctul $x = 16\ h$ și se deplasează la dreapta cu probabilitatea $p = 0{,}45$ și la stînga cu probabilitatea $q = 0{,}55$.

Conform teoriei generale expuse mai sus, probabilitatea $V(x)$ ca particula pornind din punctul x să se oprească în punctul 1, verifică ecuația diferențială

$$\frac{d^2V(x)}{dx^2} - \alpha \frac{dV(x)}{dx} = 0,$$

unde

$$\alpha = \frac{q-p}{qh} = \frac{40}{11}.$$

Prin urmare, $V(x)$ satisface ecuația diferențială (XXIII, 7.6).

Din teoria expusă mai sus, rezultă că vom găsi o valoare a soluției cu atît mai apropiată de valoarea ei exactă, cu cît pasul h este mai mic. În cazul nostru luînd $h = 0{,}05$ este evident că nu vom obține o soluție corespunzătoare din punct de vedere al exactității, cu atît mai mult cu cît și numărul de experiențe pe care-l folosim este destul de restrîns.

Deoarece $h = 0{,}05$ avem 20 de diviziuni pe intervalul (0,1). Punctele de diviziune le notăm prin 1, 2, 3, ..., 18, 19, 20. Plecăm de fiecare dată de la punctul de diviziune 16, întrucît ne-am propus să calculăm soluția ecuației diferențiale (XXIII, 7.6) in punctul $x = 0{,}8$. Probabilitatea unei deplasări la dreapta fiind 0,45 și probabilitatea unei deplasări la stînga fiind 0,55, folosind numerele întîmplătoare de la 0—100, pentru toate numerele întîmplătoare mai mari decît 55 avem o deplasare la dreapta cu o diviziune și pentru toate numerele întîmplătoare mai mici decît 55 avem o deplasare la stînga cu o diviziune. Experiența se termină cînd mersul la întîmplare ajunge la punctul 0 sau 20.

Modul cum a decurs prima experienţă, rezultă din tabelul 118.

Tabelul 118

Experienţa I

Nr. de ordine al paşilor	Numere întîmplătoare	Nr. diviziuni atinse
1	2	3
1	15	15
2	85	16
3	47	15
4	13	14
5	10	13
6	05	12
7	65	13
8	59	14
9	31	13
10	91	14
11	63	15
12	89	16
13	70	17
14	14	16
15	92	17
16	81	18
17	87	19
18	74	20

În prima coloană am dat numărul de ordine al paşilor pornind din punctul de diviziune 16. În coloana a doua sînt trecute numerele întîmplătoare. Primul număr întîmplător este 15. Acest număr este mai mic decît 55. Ca atare primul pas trebuie făcut la stînga şi cum sîntem în punctul 16, înseamnă că după primul pas am ajuns în punctul 15. Acest număr de diviziune atins după efectuarea primului pas l-am trecut în primul rînd al coloanei 3. Al doilea număr întîmplător este 85 şi deoarece este mai mare decît 55 înseamnă că din punctul 15 la care am ajuns, trebuie să facem un pas la dreapta. După al doilea pas am ajuns deci la punctul 16, pe care îl trecem în linia a doua a coloanei 3. Repetînd acelaşi procedeu, pe baza succesiunii de numere întîmplătoare din coloana 2, ne deplasăm succesiv la dreapta şi la stînga în punctele pe care le însemnăm în coloana 3. Se vede că la al 18-lea pas am atins punctul 20. Prima experienţă s-a terminat deci cu oprirea în dreapta a segmentului considerat, adică în punctul 20.

Tabelul 119

Experienţa II

Nr. de ordine al paşilor	Numere întîmplătoare	Nr. diviziuni atinse
1	2	3
1	94	17
2	69	18
3	82	19
4	25	18
5	82	19
6	83	20

Scanned with OKEN Scanner

După cum se vede (tabelul 119) și a doua experiență a condus după șase pași la o oprire la marginea din dreapta a segmentului.

Tabelul 120

Experiența III

Nr. de ordine al pașilor	Numere întîmplătoare	Nr. diviziuni atinse
1	2	3
1	73	17
2	91	18
3	33	17
4	56	18
5	88	19
6	87	20

Experiența a treia, după șase pași, ne-a condus tot la extremitatea 20.

Tabelul 121

Experiența IV

Nr. de ordine al pașilor	Nr. întîmplătoare	Nr. diviziuni atinse
1	2	3
1	32	15
2	44	14
3	94	15
4	13	14
5	78	15
6	42	14
7	04	13
8	82	14
9	38	13
10	91	14
11	54	13
12	65	14
13	49	13
14	03	12
15	56	13
16	48	12
17	23	11
18	43	10
19	40	9
20	66	10
21	77	11
22	40	10
23	69	11
24	26	10
25	55	11
26	06	10
27	50	9
28	68	10
29	31	9
30	59	10
31	59	11

Scanned with OKEN Scanner

Tabelul 121 (continuare)

Nr. de ordine al paşilor	Nr. întîmplătoare	Nr. diviziuni atinse
1	2	3
32	72	12
33	51	11
34	15	10
35	46	9
36	23	8
37	54	7
38	73	8
39	55	9
40	21	8
41	36	7
42	06	6
43	37	5
44	71	6
45	13	5
46	28	4
47	47	3
48	66	4
49	40	3
50	63	4
51	25	3
52	61	4
53	44	3
54	24	2
55	27	1
56	39	0

Experienţa a patra după 56 paşi ne-a condus de data aceasta la extremitatea stîngă a intervalului, adică în punctul 0.

A. Popescu a aplicat în continuare acest procedeu efectuînd în total 146 experienţe. Particula a ajuns în aceste experienţe de 56 de ori la diviziunea 20 şi de 90 ori la punctul 0. S-a găsit deci:

$$y \simeq \frac{56}{146} = 0{,}384.$$

Aceasta este valoarea aproximativă a soluţiei. Ea diferă de valoarea exactă 0,404 cu mai puţin de 5%, ceea ce corespunde previziunilor teoretice, deoarece intervalul $h = 0{,}05$ este mare, iar numărul probelor, egal cu 146, este destul de redus. Într-adevăr, eroarea comisă în n experienţe are valoarea

$$3\frac{\sqrt{0{,}4 \times 0{,}6}}{\sqrt{n}}$$

limitîndu-ne numai la evaluarea erorii constînd din executarea experienţelor şi lăsînd la o parte eroarea provenită din faptul că h ia o valoare finită. Prin urmare, cu o probabilitate egală cu 0,997, valoarea găsită prin aplicarea procedeului de mai sus pentru numărul de experienţe executate ($n = 146$) trebuie să fie cuprinsă între 0,282 şi 0,526. Valoarea găsită egală cu 0,384 se găseşte în intervalul teoretic de mai sus, confirmînd previziunile teoretice.

Scanned with OKEN Scanner

8. Rătăcirea întîmplătoare uniformă pe o suprafață plană

La punctul precedent ne-am ocupat de rătăcirea întîmplătoare a unei particule pe o dreaptă și am văzut că deplasările ei se încadrează într-un proces aleator, care permite rezolvarea pe cale statistică a unor ecuații diferențiale.

Vom trata aceeași problemă în plan. De data aceasta procesele aleatoare, care se nasc, conduc la ecuații cu derivate parțiale.

Acoperim planul cu o rețea de drepte echidistante, avînd distanța între ele egală cu h (vezi fig. 52).

Luăm drept axe de coordonate două drepte perpendiculare din rețea. O particulă care se găsește în nodul $M(x, y)$ poate să se deplaseze în nodurile $M_1(x, y + h)$, $M_2(x + h, y)$, $M_3(x, y - h)$, $M_4(x - h, y)$. Cu alte cuvinte la fiecare pas particula nu se poate deplasa decît în unul din nodurile vecine: în sus, în jos, la dreapta și la stînga. Probabilitatea unei deplasări pentru fiecare direcție posibilă este egală cu 0,25. Probabilitatea aceasta rămîne neschimbată, oricare ar fi poziția particulei în plan și oricare ar fi drumul parcurs de ea. Fie $P(x, y, t)$ probabilitatea ca particula ce pleacă din origină să ajungă în nodul $M(x,y)$ în timpul t. Notăm prin Δt timpul necesar efectuării unui pas. Probabilitatea $P(x, y, t)$ satisface ecuația cu diferențe finite

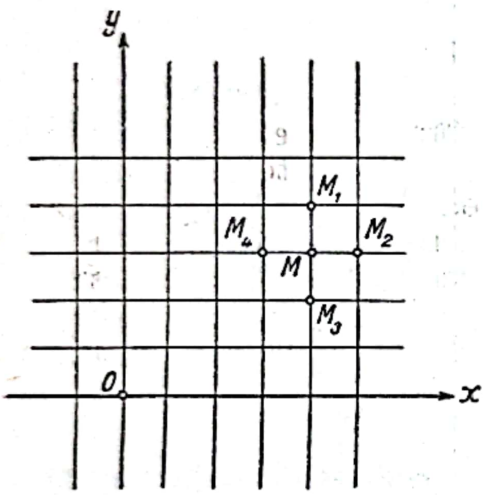


Fig. 52

$$P(x, y, t+\Delta t) = \frac{1}{4}[P(x+h, y, t) + P(x-h, y, t) + P(x, y+h, t) + P(x, y-h, t)]. \qquad \text{(XXIII, 8.1)}$$

Într-adevăr, dacă particula, plecînd din origină, ajunge în nodul $M(x, y)$ după timpul $t + \Delta t$, înseamnă că după timpul t se găsea în unul din cele patru noduri vecine, la care poate ajunge respectiv cu probabilitățile $P(x+h, y, t)$, $P(x-h, y, t)$, $P(x, y+h, t)$, $P(x, y-h, t)$. Deoarece probabilitatea deplasării de la oricare din aceste noduri la $M(x, y)$ este egală cu $\frac{1}{4}$ ajungem la relația (XXIII, 8.1).

Ecuația cu diferențe finite (XXIII, 8.1) poate fi pusă sub forma:

$$\frac{P(x, y, t+\Delta t) - P(x, y, t)}{\Delta t} = \frac{h^2}{4\Delta t}\left[\frac{P(x+h, y, t) - 2P(x, y) + P(x-h, y, t)}{h^2} + \frac{P(x, y+h, t) - 2P(x, y) + P(x, y-h, t)}{h^2}\right].$$

Făcînd pe h şi Δt să tindă către zero, astfel ca

$$\lim_{h\to 0,\Delta\to 0} \frac{h^2}{4\Delta t} = k,$$

unde k este o constantă, obţinem prin trecere la limită, din relaţia de mai sus:

$$\frac{\partial P}{\partial t} = k\left(\frac{\partial^2 P}{\partial x^2} + \frac{\partial^2 P}{\partial y^2}\right) \qquad \text{(XXIII, 8.2)}$$

adică ecuaţia căldurii.

Prin urmare procesul stohastic pe care l-am definit ne conduce la o problemă de ecuaţii cu derivate parţiale.

Mersul întîmplător poate fi cel mai simplu realizat experimental, în modul următor.

Pornind de la o tabelă cu numere întîmplătoare, pe care le transcriem în baza de numeraţie patru, un număr întîmplător va avea forma 1,0,3,2,1,3,0. El va fi scris numai cu cifrele 0,1,2,3. Vom face acum convenţia ca cifra 0 să reprezinte o deplasare la dreapta, cifra 1 o deplasare la stînga, cifra 2 o deplasare în sus şi cifra 3 o deplasare în jos (vezi fig. 52).

În modul acesta realizăm un număr mare de deplasări, plecînd din origină. Frecvenţa drumurilor care plecînd din origină se opresc în punctul $M(x,y)$ după acelaşi număr de paşi, ne dă o soluţie a ecuaţiei cu derivate parţiale (XXIII,8.1).

9. Ecuaţia lui Laplace

Ne propunem să găsim soluţia ecuaţiei lui Laplace

$$\Delta V(x, y) = \frac{\partial^2 V}{\partial x^2} + \frac{\partial^2 V}{\partial y^2} = 0 \qquad \text{(XXIII, 9.1)}$$

care pe un contur dat C, ia valori date

$$V(R) = f(R),$$

prin R notînd un punct al conturului, iar $f(R)$ fiind o funcţie continuă.

Se cunoaşte din analiza matematică că soluţia acestei ecuaţii există. Se pune problema determinării soluţiei pe baza unui proces aleator. Considerăm aceeaşi reţea ca în figura 52 şi presupunem că conturul C este format din linii ale reţelei. Considerăm acelaşi mers la întîmplare ca în p. 650. Deplasările în cele patru direcţii se fac fiecare cu probabilitatea 0,25. Pornind de la un nod M din interiorul conturului, cînd se ajunge la un punct R de pe contur mersul se termină şi se asociază $f(R)$. Pe aceeaşi cale ca la punctul precedent, se arată că valoarea medie a valorilor $f(R)$ reprezintă soluţia căutată a ecuaţiei (XXIII, 9.1).

Scanned with OKEN Scanner

Problema mersului la întîmplare, care stă la baza integrării ecuațiilor diferențiale și ecuațiilor cu derivate parțiale prin metoda mai sus indicată, a fost tratată într-o formă mai generală de mai mulți autori.

Să considerăm mai întîi mișcarea pe o axă, drept noduri luînd punctele de ordonate $x = 0, \pm 1, \pm 2 \ldots$

Notăm prin p_r probabilitatea de a trece de la punctul de abscisă r la $r+1$ și prin q_r probabilitatea de a trece de la nodul r la nodul $r-1$, $(p_r + q_r = 1)$.

Dacă

$$0 < a \leqslant p_r \leqslant b < 1,$$

oricare ar fi nodul r și dacă fiecărui nod r îi asociem o valoare $x_r > 0$, cînd o particulă prin mișcare întîmplătoare ajunge în acest nod, se poate arăta că suma numerelor x_r obținută de o particulă în mersul ei întîmplător, urmează la limită o distribuție normală *).

Proprietatea nu mai este adevărată dacă unele din probabilitățile p_r devin egale cu 1 sau cu 0.

Interesant este și mersul la întîmplare pe un tor. Să considerăm astfel pe un tor, h cercuri paralele și k cercuri meridiane. Fie (i, j) coordonatele patrulaterelor astfel formate: i variază de la $h-1$, j variază de la $k-1$. Fiecărui punct (i, j) îi corespund patru vîrfuri vecine. Să considerăm variabila x care ia valoarea a_1 în corespondență cu mișcarea $(i, j) \to (i+1, j)$ cu probabilitatea p_{ij}, valoarea a_2 pentru mișcarea $(i,j) \to (i,j+1)$ cu probabilitatea q_{ij}, valoarea a_3 pentru mișcarea $(i, j) \to (i-1, j)$ cu probabilitatea r_{ij}, valoarea a_4 pentru mișcarea $(i, j) \to (i, j-1)$ cu probabilitatea s_{ij}. Evident, avem

$$p_{ij} + q_{ij} + r_{ij} + s_{ij} = 1.$$

Să considerăm, plecînd de la punctul (i, j), succesiunea de variabile $X_1, X_2, \ldots, X_n$ corespunzătoare succesiunii de mișcări la întîmplare. Se poate arăta că în condiții destul de generale suma $X_1 + X_2 + \ldots + X_n$ urmează, cînd n crește necontenit, o distribuție normală.

Mersul la întîmplare constituie o problemă interesantă care poate conduce, după cum am văzut, la aplicații utile.

10. Aproximarea prin polinoame

S. M. Ermakov și V. G. Zolohutin, într-un recent studiu intitulat *Aproximarea prin polinoame și metoda Monte-Carlo* **) aduc o contribuție interesantă la calculul integralelor multiple prin metode statistice.

Ca variabilă aleatoare θ se ia o valoare aproximativă a integralei

$$I = \int_D f(Q)\mathrm{d}Q,$$

unde D este un domeniu din spațiul euclidian cu k dimensiuni, anume

$$\theta = \sum_{i=0}^{n} A_i f(Q_i), \qquad Q_i \in D$$

*) O. Onicescu și G. Mihoc, *Lanțuri de mișcări discontinue și schimbări de stare*. Buletinul matematic al Societății Române de științe, 48, *1—2* (1947).

**) Ermakov, Zolohutin, *Polinomialnîie priblijenia i metod Monte-Carlo*, *Teoria veroiatnostei i eio primeneniea* (1960).

care reprezintă o combinaţie liniară de valori ale funcţiei într-un număr de n puncte Q_i.

Să presupunem că $f(Q) \in L_D^2$ (spaţiul funcţiilor de pătrat sumabil pe D). Considerăm în L_D^2, $n+1$ funcţii liniar independente şi ortonormate:

$$\int_D \varphi_i(Q)\varphi_j(Q)\mathrm{d}Q = \begin{cases} 0 \text{ pentru } i \neq j, & i = 0,\ldots,n, \\ 1 \text{ pentru } i = j, & j = 0,\ldots,n. \end{cases}$$

Aceste funcţii generează un subspaţiu $(n+1)$-dimensional $L_{D,n+1}$ al lui L_D^2.

Pentru construcţia lui θ vom modifica uşor problema şi vom lua

$$\theta = \sum_{i=0}^{n} A_i f(Q_i) \approx \int_D \varphi_0(Q) f(Q)\mathrm{d}Q.$$

Formula este exactă pentru $f \in L_{D,n+1}^2$.

Ne propunem deci să calculăm valoarea integralei

$$\int_D \varphi_0(Q) f(Q)\mathrm{d}Q.$$

Dacă

$$W_{n+1}(Q_0, Q_1, \ldots, Q_n) = \det \| \varphi_0(Q_i), \ldots, \varphi_n(Q_i) \|_0^n \neq 0,$$

atunci luăm pentru θ valoarea

$$\theta = \frac{\det \| f(Q_i), \varphi_1(Q_i), \ldots, \varphi_n(Q_i) \|_0^n}{W_{n+1}(Q_0, Q_1, \ldots, Q_n)}.$$

S. M. Ermakov şi V. G. Zolohutin demonstrează două teoreme, al căror enunţ îl dăm mai jos. Cum nu ne interesează decît conţinutul acestor teoreme, în vederea aplicării lor, nu vom expune demonstraţiile, care dealminteri nu prezintă nici o dificultate.

TEOREMA 1. *Dacă $Q_0, Q_1, \ldots, Q_n$ sînt puncte întîmplătoare în spaţiul euclidian k-dimensional a căror densitate de probabilitate este*

$$F(Q_0, Q_1, \ldots, Q_n) = \frac{1}{(n+1)!} W_{n+1}^2(Q_0, Q_1, \ldots, Q_n),$$

atunci valoarea medie a variabilei θ este

$$\int_D f(Q)\varphi_0(Q)\mathrm{d}Q.$$

TEOREMA 2. *Dispersia variabilei aleatoare* θ *este*

$$D^2(\theta) = \int_D f^2(Q)\mathrm{d}Q - \sum_{i=0}^{n} \alpha_i^2,$$

unde

$$\alpha_i = \int_D f(Q)\varphi_i(Q)\mathrm{d}Q.$$

Potrivit teoremei 1 avem

$$M(\theta) = \int_D f(Q)\varphi_0(Q)\mathrm{d}Q.$$

În baza legii de probabilitate admise pentru punctele $Q_0, Q_1, \ldots, Q_n$, variabila θ se încadrează într-un proces aleator, din care rezultă valoarea ei pe cale statistică şi deci şi a integralei căutate.

Metoda expusă este o generalizare a metodei obişnuite şi se reduce la aceasta pentru $n = 0$.

Expresia lui $D^2(\theta)$ arată că cel mai convenabil este atunci cînd seria Fourier a lui $f(Q)$ în raport cu sistemul ortonormat considerat converge rapid în medie către $f(Q)$.

Teoremele 1 şi 2 arată legătura ce există între dezvoltarea în serie Fourier după un sistem ortonormat şi metoda statistică de calcul al integralelor.

Dacă, de exemplu φ_i este polinomul lui Lagrange de gradul i, $D^2(\theta)$ este integrala pătratului restului dezvoltării lui $f(Q)$ în seria Fourier după polinoamele lui Lagrange. O legătură analogă se stabileşte între interpolarea trigonometrică şi dezvoltarea în seria Fourier după funcţiile trigonometrice.

11. Inversa unei matrice

Să studiem rătăcirea întîmplătoare a unei particule, care poate trece prin poziţiile $S_1, S_2, \ldots, S_k$. Pentru particula care se află în S_j se definesc următoarele probabilităţi: a_{ij}, de deplasare din S_j în S_i şi p_j probabilitatea de a se opri în S_j. Avem

$$p_j + \sum_{i=1}^{k} a_{ij} = 1, \quad (j = 1, 2, \ldots, k)., \quad \text{(XXIII,11.1)}$$

Notînd cu B_{ij} probabilitatea ca particula ce pleacă din S_j să se oprească în S_i, se vede uşor că B_{ij} trebuie să satisfacă ecuaţiile:

$$B_{ij} = \sum_{r=1}^{k} a_{rj}B_{ir} + \delta_{ij}p_j, \quad (i,j=1,2,\ldots,k), \quad \text{(XXIII,11.2)}$$

unde $\delta_{ij} = 1$, pentru $i = j$ şi $\delta_{ij} = 0$, pentru $i \neq j$. Relaţiile (XXIII, 11.2) se pot uşor demonstra. Într-adevăr, particula ce pleacă din S_j, sau

merge în S_r cu probabilitatea a_{rj} și după aceea se oprește în S_i cu probabilitatea B_{ir}, sau trebuie să se oprească imediat cu probabilitatea p_j, în cazul cînd $i = j$.

Notînd

$$\frac{B_{ij}}{p_i} = c_{ij} \qquad \text{(XXIII, 11.3)}$$

obținem sistemul

$$c_{ij} = \sum_{r=1}^{k} a_{rj} c_{ir} + \delta_{ij}$$

sau

$$c_{ij} - \sum_{r=1}^{k} a_{rj} c_{ir} = \delta_{ij} \qquad \text{(XXIII,11.4)}$$

care se poate scrie sub forma matricială

$$\begin{Vmatrix} c_{11} & c_{12} \dots c_{1k} \\ c_{21} & c_{22} \dots c_{2k} \\ \dots & \dots \\ c_{k1} & c_{k2} \dots c_{kk} \end{Vmatrix} \cdot \begin{Vmatrix} 1-a_{11} & a_{12} & \dots & a_{1k} \\ a_{21} & 1-a_{22} \dots & & a_{2k} \\ \dots & \dots & & \dots \\ a_{k1} & a_{k2} & \dots & 1-a_{kk} \end{Vmatrix} = \begin{Vmatrix} 1 & 0 \dots 0 \\ 0 & 1 \dots 0 \\ \dots & \dots \\ 0 & 0 \dots 1 \end{Vmatrix}.$$

Notăm

$$C = \begin{Vmatrix} c_{11} & c_{12} \dots c_{1k} \\ c_{21} & c_{22} \dots c_{2k} \\ \dots & \dots \\ c_{k1} & c_{k2} \dots c_{kk} \end{Vmatrix}, \quad A = \begin{Vmatrix} a_{11} & a_{12} \dots a_{1k} \\ a_{21} & a_{22} \dots a_{2k} \\ \dots & \dots \\ a_{k1} & a_{k2} \dots a_{kk} \end{Vmatrix},$$

$$I = \begin{Vmatrix} 1 & 0 \dots 0 \\ 0 & 1 \dots 0 \\ \dots & \dots \\ 0 & 0 \dots 0 \end{Vmatrix}.$$

Produsul matricial de mai sus poate fi scris :

$$C \cdot (I - A) = I. \qquad \text{(XXIII, 11.5)}$$

Rezultă că C este inversa matricei A.

Numerele c_{ij} care sînt elementele matricei C se calculează din relațiile (XXIII, 11.2). În aceste relații probabilitățile p_i sînt cunoscute, iar probabilitățile B_{ij} rezultă statistic. Astfel, dacă din N_j drumuri pornite din S_j, N_{ij} s-au oprit în S_i, luăm pentru B_{ij} valoarea $\frac{N_{ij}}{N_j}$. Prin urmare elementele matricei inverse se calculează cu ajutorul mersului întîmplător pe care l-am indicat mai sus.

Trebuie reținut că acest procedeu nu este aplicabil decît matricelor A care au toate elementele pozitive. Dealtminteri acesta este cazul care interesează în economie, deoarece matricea legăturilor între ramuri are toate elementele pozitive.

Acest proces probabilistic se folosește și la rezolvarea sistemelor de ecuații liniare.

În acest caz la oprirea în S_i atașăm valoarea T_i, atunci speranța matematică a valorilor T_i pentru istoria particulelor care au început în S_j, pe care o notăm cu Y_i, trebuie să satisfacă ecuația

$$Y_j = \sum_{r=1}^{k} a_{rj} Y_r + p_j T_j, \quad (j = 1,2,\ldots,k). \qquad \text{(XXIII,11.6)}$$

Și acest sistem nu este rezolvabil decît în cazul cînd matricea A are toate elementele pozitive.

12. Procese cu legături complete

În diverse aplicații, întîlnim valori care trebuie să fie calculate folosind procese stohastice cu legături complete *).

Vom defini noțiunea de lanț cu legături complete cu ajutorul unui exemplu. Să presupunem că un eveniment se realizează în prima experiență cu probabilitatea p_1. În experiența a doua probabilitatea ca el să se realizeze depinde de faptul dacă s-a realizat în prima experiență sau nu. Dacă el s-a realizat în prima experiență, atunci probabilitatea p_2 ca să se realizeze în experiența a doua, este dată de operatorul liniar

$$A_1 p_1 = 0{,}3 + 0{,}5\, p_1.$$

Avem deci

$$p_2 = 0{,}3 + 0{,}5\, p_1.$$

Dacă evenimentul nu s-a realizat în experiența întîi, atunci probabilitatea ca el să se realizeze în experiența a doua se obține prin aplicarea operatorului

$$A_2 p_1 = 0{,}1 + 0{,}8\ p_1.$$

Avem în acest caz

$$p_2 = 0{,}1 + 0{,}8\ p_1.$$

Pornind acum de la probabilitatea p_2 astfel obținută, aplicînd mai departe operatorii $A_1 p$ și $A_2 p$, după cum evenimentul s-a realizat în experiența a doua sau nu, găsim probabilitatea p_3 ca evenimentul să se realizeze în experiența a treia.

*) Pentru studiul lanțurilor cu legături complete, recomandăm cartea: G. Ciucu și R. Theodorescu, *Procese cu legături complete*, Ed. Academiei R.P.R., București, 1960.

Scanned with OKEN Scanner

Aplicînd același procedeu se demonstrează că probabilitatea p_n ca evenimentul să se repete în experiențe de rangul n tinde, cînd n crește necontenit, către o limită p, independentă de valoarea inițială p_1 *).

Ne propunem să determinăm, în cazul nostru, prin metode statistice valoarea lui p, pornind de la valoarea inițială $p_1 = 0{,}40$.

Succesiunea calculelor rezultă din tabelul 122.

Tabelul 122

Nr. experienței	Nr. întîmplător	Probabilitatea evenimentului	Operator
1	2	3	4
0	0,15	0,400	A_1
1	0,85	0,500	A_2
2	0,47	0,500	A_1
3	0,13	0,525	A_1
4	0,10	0,562	A_1
5	0,05	0,581	A_1
6	0,65	0,590	A_2
7	0,59	0,572	A_2
8	0,31	0,558	A_1
9	0,91	0,579	A_2
10	0,63	0,563	A_2
11	0,89	0,550	A_2
12	0,70	0,540	A_2
13	0,14	0,532	A_1
14	0,92	0,566	A_2
15	0,81	0,553	A_2
16	0,87	0,542	A_2
17	0,74	0,534	A_2
18	0,94	0,527	A_2
19	0,69	0,522	A_2
20	0,82	0,518	A_2
21	0,25	0,514	A_1
22	0,82	0,557	A_2
23	0,83	0,546	A_2
24	0,73	0,537	A_2
25	0,91	0,530	A_2
26	0,33	0,524	A_1
27	0,56	0,562	A_1
28	0,88	0,580	A_2

În prima coloană a acestui tabel a fost trecut numărul de ordine al experienței. Coloana 2 conține numere întîmplătoare, iar în coloana 3 au fost date probabilitățile de realizare ale evenimentului în experiența respectivă. În ultima coloană a fost indicat operatorul care se aplică atunci cînd trecem de la o probă la alta. Inițial am pornit cu probabilitatea 0,40 așezată în coloana 2. Numărul întîmplător fiind 0,15, mai mic decît 0,40, înseamnă că evenimentul s-a realizat și că pentru găsirea probabilității următoare trebuie să aplicăm operatorul A_1. Acest operator aplicat la

*) O. Onicescu, G. Mihoc, *Sur les chaînes statistiques*. C.R. Acad. Sciences, Paris, 200 (1935).

Scanned with OKEN Scanner

valoarea 0,40, ne-a condus la valoarea 0,42 aşezată în coloana 3. Numărul întîmplător următor este 0,85, mai mare decît 0,42. Înseamnă că de data aceasta evenimentul nu s-a realizat şi că trebuie să aplicăm operatorul A_2. Aplicînd acest operator la valoarea 0,42, găsim valoarea 0,44 pe care o trecem în coloana 3. Aplicăm acest procedeu consecutiv. Se vede că în experienţa 28, am găsit pentru probabilitatea respectivă valoarea 0,58.

Pentru a verifica independenţa probabilităţii verificării evenimentului la experienţele de rang ridicat de valoarea probabilităţii iniţiale, refacem tabelul 122, luînd pentru probabilitatea iniţială valoarea 0,90.

Tabelul 123

Nr. experienţei	Nr. întîmplător	Probabilitatea evenimentelor	Operator
1	2	3	4
0	0,15	0,900	A_1
1	0,85	0,750	A_2
2	0,47	0,700	A_1
3	0,13	0,650	A_1
4	0,10	0,625	A_1
5	0,05	0,612	A_1
6	0,65	0,606	A_2
7	0,59	0,585	A_2
8	0,31	0,568	A_1
9	0,91	0,584	A_2
10	0,63	0,567	A_2
11	0,89	0,554	A_2
12	0,70	0,543	A_2
13	0,14	0,534	A_1
14	0,92	0,567	A_2
15	0,81	0,554	A_2
16	0,87	0,543	A_2
17	0,74	0,534	A_2
18	0,94	0,527	A_2
19	0,69	0,522	A_2
20	0,82	0,518	A_2
21	0,25	0,514	A_1
22	0,82	0,557	A_2
23	0,83	0,546	A_2
24	0,73	0,537	A_2
25	0,91	0,530	A_2
26	0,33	0,524	A_1
27	0,57	0,562	A_1
28	0,88	0,580	A_2

($p_1 = 0{,}90$; $A_1p = 0{,}3 + 0{,}5\ p$; $A_2p = 0{,}1 + 0{,}8\ p$)

Se vede că începînd de la experienţa 17 probabilităţile trecute în coloana 3 coincid, deşi probabilităţile iniţiale 0,40 şi 0,90 sînt esenţial diferite între ele. Experienţa confirmă deci, după un număr foarte mic de probe, previziunea teoretică.

Valoarea limită p se obţine făcînd media aritmetică a valorilor din coloana 3.

Scanned with OKEN Scanner

Scanned with OKEN Scanner

Capitolul XXIV

Programarea liniară

1. Considerații generale

Generalitatea operațiilor matematice a făcut posibilă aplicarea lor la studiul fenomenelor naturii și fenomenelor social-economice cele mai diverse. O relație matematică oarecare — obținută prin abstractizare și demonstrată — se aplică la realitatea concretă în tehnică, economie etc., fără să ne mai punem de fiecare dată problema exactității sau verificării ei. În modul acesta matematica a adus servicii imense științei și tehnicii moderne, al căror progres nu ar fi fost posibil fără valabilitatea generală a operațiilor matematice pe care le foloseau.

În secolul nostru — îndeosebi în ultimile decenii — asistăm însă și la un fenomen invers și anume la un impuls teoretic de seamă dat științelor matematice de dezvoltarea impetuoasă, fără precedent, a tehnicii moderne și pătrunderea ei pe scară tot mai largă în activitatea de producție. Se știe că în ultimii ani au apărut mașinile electronice de calcul capabile să facă mii de operații de calcul cu numere pe secundă. Prin punerea în funcțiune a acestor mașini a rezultat un mare progres al matematicilor aplicate. Pînă acum, soluțiile matematice ale problemelor ce se puneau în diverse ramuri ale științei și tehnicii, care aveau calcule numerice foarte numeroase, erau de multe ori abandonate din cauza imposibilității efectuării lor într-un timp util sau din cauza costului lor prea ridicat. Matematicienii înșiși au renunțat să mai abordeze asemenea metode care nu-și puteau atinge scopul practic pentru care erau create. Mașinile electronice de calcul au înlăturat această dificultate. Metode matematice care erau simple din punct de vedere teoretic, dar erau inaplicabile în practică din cauza imensității calculelor numerice la care conduceau, sînt astăzi folosite pe scară largă în diferite ramuri ale științei, inclusiv în științele social-economice. Una din aceste metode — sau mai exact — unele din aceste metode sînt cuprinse sub denumirea de *programarea liniară*.

Problemele ce se rezolvă cu ajutorul acestor metode sînt de tipul celor consacrate planificării optime, adică probleme de alegere a variantei de plan celei mai bune pentru folosirea resurselor materiale , băneşti şi de forţă de muncă ale unei întreprinderi, ramuri economice etc. Un asemenea plan optim asigură fie realizarea sarcinii propuse cu cheltuieli minime, fie depăşirea sarcinii (atingerea unui rezultat maxim posibil) cu resursele materiale şi băneşti existente *).

Importanţa rezolvării acestor probleme pentru ridicarea eficacităţii economice a activităţii de producţie în cadrul larg al construcţiei socialismului şi comunismului în ţara noastră este atît de evidentă încît nu necesită nici un fel de consideraţii doveditoare. De acest lucru sînt convinşi deopotrivă atît matematicienii cît şi economiştii şi statisticienii, care acordă atenţie tot mai mare problemelor teoretice ale programării liniare cît şi aplicaţiilor ei practice **).

Atenţie deosebită se acordă programării liniare şi în U.R.S.S., ca şi în toate ţările de democraţie populară (R. P. Polonă, R. P. Ungară, R. P. Romînă, R. S. Cehoslovacă, R.D.G. şi altele), unde au apărut studii teoretice valoroase şi aplicaţii practice ale programării liniare.

Metodologia calculului matematic al variantelor de plan optim a fost elaborată minuţios, pentru prima dată, de L. V. Kantorovici ***). Ea se deosebeşte foarte puţin de metoda programării liniare în accepţia ei modernă.

În S.U.A. şi în alte ţări occidentale ****), problemele legate de calculul programului optim de activitate operativă au fost abordate abia începînd din anul 1949, pentru cadrul îngust al întreprinderilor capitaliste, folosindu-se în acest scop unele rezultate obţinute în timpul celui de-al doilea război mondial în domeniul aplicării metodelor matematice la rezolvarea problemelor operaţionale şi de strategie militară.

În condiţiile economiei capitaliste, în care domină proprietatea particulară asupra mijloacelor de producţie şi deci este cu neputinţă planificarea economiei naţionale, aplicarea programării liniare pe ramuri sau pe ansamblul economiei naţionale este evident imposibilă, şi de aceea, însăşi perspectivele dezvoltării acestei metode sînt înguste, limitate.

---

*) Noţiunea de „resurse" este folosită aici şi mai departe în sensul ei cel mai larg, cuprinzînd în afară de materii prime şi materiale şi diferite categorii de energie, forţă de muncă, unelte de producţie etc.

**) În ultimul timp în revistele de specialitate din ţara noastră au apărut numeroase studii teoretice şi practice ca şi o serie de comunicări pe tema programării liniare ţinute în cadrul colocviilor de calcul numeric organizate de Academia R.P.R. şi filialele ei (Iaşi, Cluj); de asemenea, în diferite departamente se fac aplicaţii practice utile de programări liniare.

***) L. V. Kantorovici, *Matematiceskie metodi organizaţii i planirovania proizvodstva*. L. G. U. Leningrad, 1939. În această lucrare metoda cunoscută azi sub denumirea „programarea liniară", poartă numele de „metoda factorilor rezolvanţi" / în limba rusă: „Metod razresaiuşcih mnojitelei".

****) În iulie 1949, la Chicago, a avut loc o consfătuire a economiştilor şi statisticienilor în cadrul căreia s-au susţinut peste 20 de referate pe tema programării liniare, care în 1951 au apărut într-o culegere intitulată *Activity Analysis of Production and Allocation*, John Willey and Sons, New York, 1951. Această culegere constituie una din lucrările principale pe tema programării liniare. Mai tîrziu au apărut şi o serie de alte lucrări pe aceeaşi temă în S.U.A., Anglia, Norvegia şi alte ţări capitaliste.

Scanned with OKEN Scanner

Programarea liniară își găsește un cîmp de aplicare foarte larg în condițiile economiei socialiste, incomparabil mai larg decît în condițiile capitalismului. Într-o comunicare prezentată de I. P. Gherciuk la Casa oamenilor de știință din Moscova a Academiei de științe a U.R.S.S. *) arată că aplicarea judicioasă a programării liniare poate îmbogăți considerabil utilizarea tuturor resurselor economiei naționale. În special în planificare: la nivelul întreprinderii, ramurii sau al economiei naționale; la elaborarea planurilor de transport ale economiei naționale sau regiunilor economice; la repartizarea mijloacelor de mecanizare în construcții pe diferite feluri de lucrări; la alcătuirea planurilor de folosire a materiei prime complexe și la formarea amestecurilor, avînd proprietăți fixate de diferite componente (rații de furajare a animalelor etc.); la alcătuirea asolamentelor raționale; la repartizarea optimă a resurselor de muncă pe ramuri în marile gospodării de stat și colective etc. — în toate aceste probleme și în multe altele, programarea liniară se poate aplica cu deosebit succes. Eficiența aplicării ei se vădește și mai mult în analiza legăturilor de balanță între diferite ramuri ale economiei naționale ca și pe plan teritorial, interregional.

Programarea liniară este o metodă științifică de rezolvare, prin mijloace matematice, a problemelor de tipul celor arătate mai sus, găsindu-se soluțiile lor optime.

În esență — privită din punct de vedere matematic — *programarea liniară constă în determinarea valorii maxime sau minime a unei funcții statistice liniare de mai multe variabile*. În procesul acestei determinări se ține seama de latura economică a programării liniare, funcțiile respective fiind supuse unor anumite condiții, decurgînd din esența calitativă a fenomenelor economice considerate.

Să ne oprim asupra unor exemple de probleme simple ce se rezolvă de obicei, prin aplicarea programării liniare, subliniind cu acest prilej, trăsăturile specifice ale acestei metode:

1°. Avem un parc determinat de mașini-unelte și ne este cunoscut sau planificat timpul de lucru al fiecărei mașini în parte. Ne este dat, de asemenea, numărul de piese care poate fi executat cu ajutorul acestor mașini-unelte. Problema care se pune este aceea de a se întocmi un *plan optim de încărcare* a mașinilor-unelte, adică, de a folosi timpul de prelucrare a fiecărei piese, la fiecare mașină în parte în așa mod, ca timpul total de lucru pentru producerea tuturor pieselor să fie minim, sau — ceea ce este echivalent — ca producția totală pe o unitate determinată de timp să fie maximă. Varianta de distribuție a timpului de lucru pe fiecare mașină-unealtă care să satisfacă această condiție, va fi tocmai varianta de încărcare optimă căutată. Se înțelege că în acest caz, trebuie să ținem seama și de *normele de timp de lucru* necesar pentru prelucrarea fiecărei piese, la fiecare mașină-unealtă în parte.

2°. Ni se cere să formăm un *amestec* care să necesite cel mai mic volum de cheltuieli, cunoscînd că elementele componente ale acestui

*) Vezi Vestnik Statistiki, 87—88 3 (1958) — o informare asupra acestei comunicări, intitulată: *Programarea liniară — o nouă metodă de găsire a soluțiilor optime ale problemelor economice.*

amestec participă la formarea lui în proporţii diferite şi că preţul lor este diferit. Probleme de acest tip sînt numeroase. De exemplu, problema producerii unui furaj cu anumite calităţi nutritive sau problema amestecului de benzină de calităţi diferite care să satisfacă anumite condiţii tehnice dinainte stabilite şi totdeodată — să necesite cheltuieli minime; şi alte asemenea probleme.

3°. Ni se cere să transportăm o anumită cantitate de marfă de la 5 unităţi de producţie la 10 centre de depozitare, cunoscînd cantităţile ce se află asupra fiecăreia din cele 5 unităţi de producţie şi cantităţile care trebuie transportate la cele 10 centre de depozitare; de asemenea, se cunoaşte timpul necesar, în zile, pentru transportul unei încărcături sau preţul de cost unitar de la fiecare centru de depozitare la fiecare unitate de producţie. Ca şi în exemplele precedente, problema care se pune este aceea de a se stabili un *plan de transport optim* — plan care să satisfacă simultan exigenţele de aprovizionare ale tuturor celor 10 centre de depozitare, în funcţie de cantităţile aflate asupra celor 5 unităţi de producţie şi totodată, să fie cel mai economic (să necesite în total minimul de zile sau minimul de cheltuieli pentru efectuarea tuturor operaţiilor de transport).

Interesul practic pe care îl prezintă aflarea soluţiilor în problemele de tipul celor de mai sus este clar. În toate cele trei exemple de probleme de programare liniară date mai sus se remarcă unele caracteristici tipice şi anume:

a) Se pune problema găsirii unei soluţii *optime* din punct de vedere economic, în condiţiile concrete ale problemei date şi care pe plan teoretic, matematic, reprezintă „maximul" sau „minimul" unei funcţii de mai multe variabile. De obicei mulţimea variabilelor dau naştere la un sistem de ecuaţii sau inecuaţii de gradul întîi cu foarte multe necunoscute. De cele mai multe ori, numărul necunoscutelor este cu mult mai mare decît numărul ecuaţiilor. Toate acestea fac ca problema aflării soluţiei optime să devină dificilă, depăşind, din acest punct de vedere, calculul elementar al minimului sau maximului unei funcţii obişnuite.

b) Problemele de programare liniară reprezintă *modele complexe de mai mulţi factori interdependenţi*. Caracteristica fundamentală a acestor modele constă nu numai în faptul că valorile optime trebuie găsite pentru mai multe variabile, ci şi în aceea că şi aceste variabile sînt *interdependente*, condiţionîndu-se reciproc: dacă se schimbă valoarea uneia dintre aceste variabile se schimbă nu numai rezultatul care depinde de ea ci şi valorile posibile ale altor variabile din sistem. Aşa, de pildă, dacă — la căutarea variantei optime a planului de încărcare a utilajului (vezi prima problemă de mai sus) — stabilim pentru una din maşini o anumită cantitate de piese ce trebuie prelucrate, prin aceasta stabilim simultan, nu numai timpul de lucru necesar pentru executarea pieselor respective la această maşină, ci şi o anumită condiţie restrictivă de utilizare a celorlalte maşini-unelte.

Esenţialul în toate problemele de programare liniară constă în aceea că *întreaga colectivitate de resurse materiale* de care dispune *fiecare unitate economică* să fie folosite cu eficacitatea economică maximă din punctul de vedere al sarcinii de rezolvat. Se ridică, însă, imediat, o altă problemă. Resursele aflate la dispoziţie sînt de obicei în cantităţi limitate şi de cali-

Scanned with OKEN Scanner

tăţi diferite şi de aceea se naşte necesitatea de a fi economisite, de a fi raţional folosite şi în general, necesitatea unui regim care să asigure eficacitatea economică maximă în folosirea acestor resurse. Din această restricţie de folosire a resurselor derivă şi condiţionarea reciprocă pe linia afectării lor unor anumite destinaţii. Dacă o anumită parte a resurselor de care dispunem o afectăm unui anumit scop, pentru a obţine un rezultat bine determinat, prin acest fapt ne limităm de a afecta unui alt scop oarecare, numai partea de resurse rămasă, avînd alte caracteristici sub aspectul eficacităţii.

Rezultă, deci, că ori de cîte ori luăm măsuri pentru a destina unui scop oarecare o anumită resursă de care dispunem şi aceasta, evident, cu scopul de a obţine un rezultat bine determinat — noi acţionăm nu numai direct asupra factorilor care determină acest rezultat ci şi indirect, prin corelaţia de sisteme dintre aceşti factori, deosebiţi prin particularităţile calitative ale resurselor angajate.

Exemplele şi observaţiile de mai sus se găsesc, în general, în diferite lucrări de specialitate.

Vrem să subliniem aici şi consideraţiile interesante făcute pe marginea lor de I. P. Gherciuk *), indicînd criteriile esenţiale de interpretare şi înţelegere a problemelor de programare liniară.

Pentru înţelegerea corectă a esenţei problemelor de programare liniară — arată acest autor — este important să avem în vedere nu numai faptul că ele sînt probleme de determinare a unor valori optime pentru o funcţie de mai multe variabile; prin aceasta încă nu este relevată esenţa acestor probleme. Este cu mult mai important să nu pierdem din vedere că în toate problemele de programare liniară avem de-a face cu colectivităţi bine determinate de factori elementari pe care trebuie să-i folosim cu cea mai mare eficacitate şi că sîntem totodată, riguros impuşi la o serie de condiţii restrictive ce decurg din legătura de sistem dintre aceşti factori şi care caracterizează natura lor economică.

Prin urmare, sarcina programării liniare va consta în a *combina* în chipul cel mai raţional posibil, *toţi factorii elementari*, dîndu-le destinaţii care să asigure eficacitatea maximă a *întregului ansamblu de factori legaţi în sistem*.

Abordarea din acest punct de vedere a problemelor de programare liniară ne poate releva, pe de o parte, natura lor economică, iar pe de altă parte — sensul lor matematic ca şi conţinutul real al interdependenţei variabilelor din sistem.

Pentru rezolvarea problemelor de programare liniară s-au elaborat mai multe metode sau procedee tehnice de calcul, găsindu-se cu uşurinţă soluţiile optime chiar şi în cazurile cînd este considerat un mare număr de variabile interdependente. Cu alte cuvinte, tehnica modernă de determinare a valorii maxime sau minime a unei funcţii liniare de mai multe variabile a progresat mult în comparaţie cu metodele clasice de rezolvare a unor asemenea probleme.

*) Vezi articolul acestui autor din broşura *Primenenie lineinogo programirovania za-granitei*. Academia de ştiinţe a U.R.S.S. (filiala pentru cercetările ştiinţifice în Siberia), Moscova, 1959.

Problema programării liniare. Una din metodele care au cele mai mari perspective de dezvoltare, care este mai uşor de aplicat şi a cărei problemă-tip se întîlneşte mai des este programarea liniară. Exemple de probleme care se pot rezolva cu ajutorul programării liniare se pot da din domeniile cele mai diferite :

I. O întreprindere dispune de anumite resurse : materii prime, maşini, forţe de muncă etc. Ea poate fabrica diferite produse întrebuinţînd resursele cantitativ diferit pe unitate de produs. Se presupun cunoscute cantităţile disponibile din fiecare resursă, preţul unităţii fiecărui produs şi cantitatea din fiecare resursă necesară fabricării unităţii de produs. Se cere să se determine ce produse trebuie fabricate şi în ce cantităţi pentru a se obţine o valoare totală maximă a producţiei întreprinderii.

Vom nota :

m, numărul resurselor ;

n, numărul produselor ;

b_i, numărul maxim de unităţi disponibile din resursa i ;

a_{ij}, numărul unităţilor din resursa i necesare pentru fabricarea unei unităţi din produsul j ;

c_j, valoarea unitară a produsului j ;

x_j, numărul unităţilor din produsul j necesar pentru a face maximă valoarea produselor fabricate.

În formulare matematică : va trebui să determinăm cantităţile x_j aşa fel încît valoarea totală a produselor fabricate $\Sigma c_j x_j$ să fie maximă şi totodată în aşa fel încît numărul de unităţi din resursa i, necesar pentru fabricarea tuturor produselor $\Sigma a_{ij} x_j \leqslant b_i$, adică să fie mai mic sau cel mult egal cu numărul maxim de unităţi disponibile din resursa i. Ţinînd seama de faptul că x_j este un număr de unităţi, reiese condiţia $x_j \geqslant 0$, $(j = 1, 2, \ldots, n)$. Deci :

Să se determine $x_1, \ldots, x_n$ aşa fel încît

$$
\begin{aligned}
& x_1 \geqslant 0,\ x_2 \geqslant 0,\ \ldots,\ x_n \geqslant 0, \\
& a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n \leqslant b_1, \\
& a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n \leqslant b_2, \\
& \ldots\ldots\ldots\ldots\ldots\ldots\ldots \\
& a_{m1}x_1 + a_{m2}x_2 + \ldots + a_{mn}x_n \leqslant b_m, \\
& c_1x_1 + \ldots + c_nx_n = \text{maxim.}
\end{aligned}
$$

II. O problemă absolut analogă se poate ivi în agricultură. Puteam avea şi aici un număr de m resurse (pămînt, seminţe, forţe de muncă etc) cu care se pot face n culturi (grîu, porumb, sfeclă de zahăr etc.). Ştiind că avem disponibile un număr de b_i unităţi din resursa i, că sînt necesare a_{ij} unităţi din resursa i pentru cultura j şi că c_j este costul întregului proces de cultivare a culturii j la hectar, să se determine x_j, $(j=1, 2, \ldots n)$, cantităţile de resurse necesare pentru a obţine venitul maxim.

III. Problema nutriţiei sau a regimului alimentar. Un om are posibilitatea să se hrănească cu un număr de n alimente (pîine, lapte, carne etc.). Ele conţin m elemente nutritive (proteină, glucoză etc.). Omul are nevoie

de o cantitate de cel puțin b_i unități din elementul nutritiv i. Știind că în compoziția unei unități din alimentul j intră a_{ij} unități din elementul nutritiv i și că c_j este costul unei unități din alimentul j, să se determine ce cantitate din fiecare aliment trebuie să consume omul pentru a ajunge la regimul alimentar cel mai economic.

Avem în primul rînd, $x_j \geqslant 0$ prin însăși natura datelor. Obiectivul problemei este de a minimiza costul total al alimentelor $\sum_{j=1}^{n} c_j x_j$, fără ca regimul alimentar să fie nesatisfăcător, adică toate alimentele să conțină cel puțin minimul b_i din alimentul nutritiv i :

$$\Sigma a_{ij} x_j \geqslant b_i ;$$

sau :

$$\begin{aligned} & x_1 \geqslant 0, \ x_2 \geqslant 0, \ \ldots, \ x_n \geqslant 0, \\ & a_{11} x_1 + a_{12} x_2 + \ldots + a_{1n} x_n \geqslant b_1, \\ & \ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ & a_{m1} x_1 + a_{m2} x_2 + \ldots + a_{mn} x_n \geqslant b_m, \\ & c_1 x_1 + c_2 x_2 + \ldots + c_n x_n = \text{minim}. \end{aligned}$$

Exemple de astfel de tipuri de probleme se mai pot da multe. Vom mai aminti problema amestecurilor și problema transporturilor. Ele se scriu matematic pînă la urmă, sub aceeași formă ca la problema I sau la problema III. Se observă imediat că una este o problemă de minim cu inegalități „$\geqslant$" și alta de maxim, cu inegalități „$\leqslant$". Ele nu diferă însă esențial din punct de vedere matematic, deoarece condițiile $x_j = 0$ rămînînd aceleași, putem transforma, de exemplu, maximul în minim schimbînd semnul coeficienților :

$$\max_{x_j} \left(\sum_{j=1}^{n} c_j x_j\right) = - \min_{x_j} \left(\sum_{j=1}^{n} - c_j x_j\right).$$

La fel inegalitățile

$$\sum_{j=1}^{n} a_{ij} x_j \geqslant b_i$$

se pot scrie

$$\sum_{j-1}^{n} (- a_{ij})\, x_j \leqslant -b_j.$$

Deci, e vorba, în fond, de o aceeași problemă matematică.

2. Metoda Simplex

Metoda Simplex a fost concepută de G. B. Dantzig. Ea poate fi folosită la rezolvarea tuturor problemelor de programare liniară. În plus, la aplicarea acestei metode pot fi utilizate mașinile electronice de calcul, ceea ce constituie un mare avantaj cînd datele problemei sînt numeroase.

Vom arăta modul de aplicare al acestei metode pe un exemplu, indicînd modul de succesiune a operațiilor. Regulile de aplicare ale metodei sînt simple și pot fi ușor reținute. Justificarea matematică a regulilor este

expusă mai departe, însă poate fi lăsată la o parte de cititorul care urmărește numai aplicațiile practice.

Să presupunem că o gospodărie colectivă dispune de o suprafață de 900 ha, de o sumă de 2 500 000 lei și de 55 000 om-zile de muncă. Ea poate folosi următoarele culturi:

Cultura V_1 la care are nevoie la ha de 2 000 lei și 40 om-zile muncă;

Cultura V_2 la care are nevoie la ha de 3 000 lei și 60 om-zile de muncă;

Cultura V_3 la care are nevoie la ha de 2 500 lei și 60 om-zile de muncă;

Cultura V_4 la care are nevoie la ha de 2 400 lei și 50 om-zile de muncă.

Venitul net la ha este 1 600 lei pentru V_1, 3 840 lei pentru V_2, 3 400 lei pentru V_3 și 2 310 lei pentru V_4.

Se cere un plan optim de folosire al acestor culturi.

Se vede imediat că cea mai bună soluție ar fi aplicarea culturii V_2 pe toate cele 900 ha. Aceasta ar implica însă o cheltuiala de 2 700 000 lei, depășind suma disponibilă de 2 500 000 lei. Prin urmare trebuie găsită o soluție optimă, care să rezolve problema în limitele sumelor disponibile de bani și om-zile de muncă.

Principiul de bază al metodei Simplex constă în considerarea unei situații posibile de aplicare a culturilor și în îmbunătățirea acestei situații în sensul de a obține o soluție mai avantajoasă. Este evident că aplicarea succesivă a acestui principiu ne conduce la soluția optimă.

Modul de aplicare al metodei Simplex rezultă din tabelul 124.

Tabelul 124

	t_i	Baza	—	1 600	3 840	3 400	2 310	0	0	0
			V_0	V_1	V_2	V_3	V_4	V_5	V_6	V_7
I	0	V_5	900	1	1	1	1	1	0	0
	0	V_6	2 500 000	2 000	3 000	2 500	2 400	0	1	0
	0	V_7	55 000	40	60	60	50	0	0	1
	s_i			0	0	0	0	0	0	0
	$s_i - t_i$			−1 600	−3 840	−3 400	−2 310			
II	0	V_5	200/3	1/3	0	1/6	1/5	1	−1/3 000	0
	3 840	V_2	2 500/3	2/3	1	5/6	4/5	0	1/3 000	0
	0	V_7	5 000	0	0	10	2	0	−1/50	1
	s_i			2 560	3 840	3 200	3 072	0	1,28	
	$s_i - t_i$			960	0	−200	762	0	1,28	
III	3 400	V_3	400	2	0	1	6/5	6	−1/500	0
	3 840	V_2	500	−1	1	0	−1/5	−5	7/3 000	0
	0	V_7	1 000	−20	0	0	−10	−60	0	1
	s_i			2 960	3 840	3 400	3 312	1 200	2,16	
	$s_i - t_i$			1 360	0	0	1 002	1 200	2,16	

Scanned with OKEN Scanner

În acest tabel, pe lîngă culturile V_1, V_2, V_3, V_4 am trecut culturile V_5, V_6, V_7. Dacă întrebuinţăm o notaţie vectorială, putem scrie:

$$V_1 = \begin{pmatrix} 1 \\ 2\,000 \\ 40 \end{pmatrix}, \quad V_2 = \begin{pmatrix} 1 \\ 3\,000 \\ 60 \end{pmatrix}, \quad V_3 = \begin{pmatrix} 1 \\ 2500 \\ 60 \end{pmatrix}, \quad V_4 = \begin{pmatrix} 1 \\ 2\,400 \\ 50 \end{pmatrix},$$

$$V_5 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \quad V_6 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \quad V_7 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}.$$

În plus mai introducem vectorul:

$$V_0 = \begin{pmatrix} 900 \\ 2\,500\,000 \\ 55\,000 \end{pmatrix}.$$

Prima componentă a vectorului indică aria cultivată, a doua cheltuială în lei necesară pentru 1 ha şi a treia numărul de om-zile muncă pentru 1 ha.

Vectorii V_5, V_6, V_7 sînt vectorii unitate, baza în care scriem ceilalţi vectori. Ei pot fi interpretaţi ca şi vectorii V_1, V_2, V_3, V_4. De exemplu, vectorul V_5 poate fi considerat ca o cultură căreia i-am destinat o arie de 1 ha, însă 0 lei şi 0 om-zile muncă. Este evident că în astfel de condiţii nu se poate lucra şi că V_5 reprezintă o cultură care ne dă venitul 0. La fel culturile V_6, V_7 aduc venitul net 0.

Vectorul V_0 reprezintă valorile disponibile. El poate fi scris:

$$V_0 = 900\ V_5 + 2\,500\,000\ V_6 + 55\,000\ V_7,$$

luînd vectorii V_5, V_6, V_7 ca bază.

Aceasta înseamnă că cele 900 ha le-am destinat culturii V_5, cei 2 500 000 lei culturii V_6 şi 55 000 om-zile muncă culturii V_7. Procedînd astfel obţinem desigur venitul zero. Dar plecînd de la această situaţie, vom ajunge la alta, la care venitul e mai mare.

În partea I a tabelului sînt trecuţi toţi vectorii V_0, V_1, V_2, V_3, V_4, V_5, V_6, V_7 cu componentele lor. În prima linie am trecut veniturile t_i pentru fiecare cultură, reprezentată printr-un vector. În prima coloană am trecut valorile t_i corespunzătoare vectorilor bază, adică V_5, V_6, V_7. În linia notată prin s_i, am trecut pentru fiecare vector V_i, $(i = 1,2, \ldots, 7)$, suma obţinută prin adunarea componentelor lui înmulţite respectiv cu valorile din prima coloană.

Ultima linie ne arată valorile $s_i - t_i$. Constatăm că cea mai mică valoare trecută în această linie este $-3\,840$ corespunzătoare vectorului V_2. Aceasta constituie un indiciu că vectorul V_2 trebuie să intre la bază, luînd locul unuia din vectorii V_5, V_6, V_7. Se pune întrebarea: al cărui loc îl va lua V_2? Pentru aceasta împărţim elementele coloanei V_0 respectiv la elementele coloanei V_2, adică considerăm numerele: (900 :1), (2 500 000 : 3 000), (55 000 :60). Constatăm că cel mai mic dintre aceste numere este

(2 500 000 :3 000) care corespunde liniei a doua în dreptul căreia este scris vectorul V_6. Deci vectorul V_2 va lua locul lui V_6.

În partea a II-a a tabelului Simplex avem vectorii de bază V_5, V_2, V_7, pe care îi trecem în coloana a doua. În coloana întîi trecem valorile t_i corespunzătoare vectorilor de bază V_5, V_2, V_7. În partea I-a a tabelului vectorul V_6 a avut componentele (0,1,0). Rezultă că în partea a II-a a tabelului, vectorul V_2, care îi ia locul, va avea aceleaşi componente (0,1,0). Pentru a obţine acest lucru, împărţim toate elementele din linia V_6 a părţii I-a, prin 3 000 şi le trecem în linia corespunzătoare vectorului V_2 din partea a II-a.

Din linia V_5 din partea I-a, scădem elementele liniei V_2 din partea a II-a şi rezultatele le trecem la elementele liniei V_5 din partea a II-a.

Înmulţim elementele liniei V_2 din partea a II-a cu 60 şi le scădem din elementele liniei V_7 din partea I-a. Rezultatele le trecem în linia V_7 din partea a II-a a tabelului.

Linia s_i din partea a II-a se obţine prin adunarea elementelor din prima coloană înmulţite respectiv cu componentele vectorilor V_i, ($i = 1, 2, 3, \ldots, 7$). Ultima linie rezultă imediat scăzînd din valorile s_i valorile t_i corespunzătoare.

În baza părţii a II-a a tabelului Simplex, vectorul V_0 poate fi scris

$$V_0 = \frac{200}{3} V_5 + \frac{2\,500}{3} V_2 + 5\,000\, V_7;$$

în baza acestei formule înseamnă că am destinat $\frac{2\,500}{3}$ ha pentru cultura V_2, iar restul hectarelor pînă la 900 ha, pentru culturile V_5 şi V_6 care nu produc nimic. Cele $\frac{2\,500}{3}$ ha destinate culturii V_2 ne dau însă un venit net egal cu :

$$\frac{2\,500}{3} \times 3\,840 = -\frac{2\,500}{3}(s_2 - t_2). \qquad \text{(XXIV,2.1)}$$

Faţă de venitul 0 din partea I-a a tabelului am ajuns la venitul (XXIV, 2.1). Observăm însă că o îmbunătăţire a situaţiei nu poate exista decît dacă expresia (XXIV,2.1) este pozitivă sau $s_3 - t_3 < 0$. Dacă toate valorile $s_i - t_i$ ar fi fost pozitive, se vede din examinarea formulei (XXIV, 2.1) că nu am fi ajuns la o sumă pozitivă, cu care să crească venitul. Prin urmare pentru ca să putem obţine o situaţie mai bună, este necesar ca dintre valorile $s_i - t_i$ cel puţin una dintre ele să fie negativă. Dacă toate valorile $s_i - t_i$ sînt pozitive, înseamnă că nu putem ajunge la o situaţie mai bună, ceea ce caracterizează soluţia optimă.

Observăm la partea a II-a a tabelului Simplex că avem o singură valoare $s_i - t_i$ negativă şi anume $s_3 - t_3 = -200$. Putem deci, prin acelaşi procedeu, să ajungem la o situaţie mai avantajoasă. Trebuie să introducem vectorul V_3 în bază. Pentru a afla locul căreia din vectorii V_5, V_2, V_7 îl va lua, împărţim componentele vectorului V_0 respectiv la componentele

vectorului V_3 şi obţinem numerele : $\left(\frac{200}{3} : \frac{1}{6}\right)$, $\left(\frac{2\,500}{3} : \frac{5}{6}\right)$, $(5\,000 : 10)$. Dintre aceste numere primul este cel mai mic. Prin urmare vectorul V_3 va lua locul vectorului V_5.

În consecinţă în coloana a doua a părţii a III-a vom trece vectorii de bază V_3, V_2, V_7, iar în prima coloană valorile t_i corespunzătoare acestor vectori. Linia V_3 din partea a III-a se obţine împărţind elementele liniei V_5 din partea a II-a cu $\frac{1}{6}$. Din elementele liniei V_2 din partea a II-a scădem elementele liniei V_3 din partea a III-a înmulţite cu $\frac{5}{6}$, şi rezultatele le trecem în linia V_2 din partea a III-a. Din elementele liniei V_7 din partea a II-a scădem elementele liniei V_3 înmulţite cu 10.

Rezultatele astfel obţinute sînt trecute în linia V_7 din partea a III-a.

Completăm partea a III-a înmulţind elementele primei coloane cu componentele vectorilor V_i, $(i = 1, 2, \ldots, 7)$ şi însumînd rezultatele rezultă linia s_i. Am obţinut pentru vectorul V_0 soluţia :

$$V_0 = 500\ V_2 + 400\ V_3 + 1000\ V_7. \qquad \text{(XXIV,2.2)}$$

Observăm că nici una din valorile $s_i - t_i$ din partea a III-a nu este negativă. Aceasta înseamnă că mersul operaţiilor trebuie oprit şi că (XXIV, 2.2) ne dă soluţia optimă.

Prin urmare pentru a obţine un venit maximum trebuie să se folosească 500 ha din cultura V_2 şi 400 ha din cultura V_3.

Venitul obţinut este

$$500 \times 3\,840 + 400 \times 3\,400 = 3\,280\,000 \text{ lei};$$

orice alt plan de cultură, în condiţiile fixate, produce un venit mai mic.

Indicaţiile pe care le-am dat arată modul de aplicare al metodei Simplex. La p. 674, cititorul găseşte expusă justificarea matematică a aplicării acestor operaţii.

3. Bazele matematice ale programării liniare

Vom expune, pe scurt, cîteva consideraţii matematice, care justifică procedeele pe care le-am indicat în rezolvarea practică a programării liniare.

Problema generală a programării liniare se poate enunţa în modul următor :

Să se determine vectorul $(x_1, x_2, \ldots, x_n)$ care face maximă forma liniară

$$c_1x_1 + c_2x_2 + \ldots + c_n x_n \qquad \text{(XXIV, 3.1)}$$

și care satisface, în același timp, condițiile

$$\begin{aligned}
&a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n = b_1\\
&a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n = b_2,\\
&\ldots\ldots\ldots\ldots\ldots\ldots\\
&a_{m1}x_1 + a_{m2}x_2 + \ldots + a_{mn}x_n = b_m,\\
&x_1 \qquad\qquad\qquad\qquad \geqslant 0,\\
&\quad x_2 \qquad\qquad\qquad\qquad \geqslant 0,\\
&\qquad\cdot\\
&\qquad\cdot\\
&\qquad\cdot\\
&\qquad x_n \qquad\qquad\qquad\qquad \geqslant 0,
\end{aligned}$$

unde a_{ij}, b_i și c_j, $(i = 1,2, \ldots, m)$, $(j = 1,2, \ldots, n)$, $m < n$ sînt niște constante date.

Folosind notația matricială, problema generală de programare liniară se poate scrie mai simplu astfel:

$$cX = max, \qquad \text{(XXIV, 3.2)}$$

$$AX = b, \qquad \text{(XXIV, 3.3)}$$

$$X \geqslant 0, \qquad \text{(XXIV, 3.4)}$$

unde A este matricea (a_{ij}), iar b și X respectiv vectorii coloană

$$\begin{pmatrix} b_1\\ b_2\\ \cdot\\ \cdot\\ b_m \end{pmatrix}, \quad \begin{pmatrix} x_1\\ x_2\\ \cdot\\ \cdot\\ x_n \end{pmatrix}$$

iar c vectorul linie $(c_1, c_2, \ldots, c_n)$.

Forma liniară cX poartă numele de *obiectivul* sau *funcția obiectiv* a problemei. O *soluție posibilă* a unei probleme de programare liniară este un vector X care satisface simultan relațiile (XXIV,3.3) și (XXIV,3.4). O *soluție optimă* a aceleiași probleme este orice soluție posibilă care face maximă (minimă) funcția obiectiv a problemei, adică un vector X care satisface simultan relațiile (XXIV,3.2), (XXIV,3.3), (XXIV,3.4). Mulțimea soluțiilor posibile X formează *spațiul soluțiilor* problemei de programare liniară.

Considerăm ca exemplu sistemul de ecuații și funcția obiectiv pentru cazul a două necunoscute x_1 și x_2:

$$\begin{aligned}
&x_1 \qquad\qquad \geqslant 0,\\
&\quad x_2 \qquad\qquad \geqslant 0,\\
&a_{11}x_1 + a_{12}x_2 = b_1,\\
&a_{21}x_1 + a_{22}x_2 = b_2,\\
&c_1x_1 + c_2x_2 = \max.
\end{aligned}$$

O funcție liniară X de punctele $x_1, x_2, \ldots, x_m$ din spațiul euclidian R_n:

$$x = \alpha_1 x_1 + \alpha_2 x_2 + \ldots + \alpha_m x_m,$$

unde α_i, $(i = 1,2, \ldots, m)$, sînt niște scalari $(\alpha_i \geqslant 0, \sum_{i=1}^{m} \alpha_i = 1)$ se numește o *combinație convexă* de punctele x_i, $(i = 1,2, \ldots, m)$, și este evident tot un punct. O mulțime de puncte din spațiul R_n este *convexă* dacă, pentru orice pereche de puncte x_1 și x_2 aparținînd acelei mulțimi, orice combinație convexă

$$x = \alpha_1 x_1 + \alpha_2 x_2$$

aparține de asemenea mulțimii considerate. Orice punct de pe un segment care unește două puncte din spațiul R_n este o combinație convexă de cele două puncte. Reciproc, orice punct care este o combinație convexă a două puncte din spațiul R_n se găsește pe segmentul care unește cele două puncte. Rezultă, deci, că din punct de vedere geometric, o mulțime convexă este acea mulțime de puncte care conține toate segmentele care unesc două puncte oarecare aparținînd acestei mulțimi de puncte. Întreg spațiul R_n este un exemplu de mulțime convexă. De asemenea un cerc, un cub, sînt exemple de mulțimi convexe în spațiul R_2, respectiv R_3. O mulțime care nu conține decît un singur punct este, evident, o mulțime convexă. Mulțimea punctelor care se găsesc pe frontiera unui cerc, de pildă, nu este o mulțime convexă.

Un *punct extrem* al unei mulțimi convexe este un punct aparținînd mulțimii convexe, care nu este o combinație convexă de alte două puncte distincte ale mulțimii convexe. Din punct de vedere geometric înseamnă că, un punct extrem al unei mulțimi convexe este un punct al acestei mulțimi care nu se găsește pe nici un segment ce unește două puncte oarecare din aceeași mulțime distincte de primul. Astfel, fiecare punct de pe frontiera unui cerc este un punct extrem al mulțimii convexe, formată din mulțimea punctelor aparținînd atît frontierei, cît și interiorului cercului. Vîrfurile unui triunghi sînt punctele extreme ale acestei mulțimi convexe.

La baza rezolvării unei probleme de programare liniară stau următoarele două teoreme fundamentale:

TEOREMA I. *Mulțimea tuturor soluțiilor posibile ale unei probleme de programare liniară este o mulțime convexă în spațiul soluțiilor.*

Teorema este evidentă, dacă problema de programare liniară are o singură soluție. Să presupunem acum că problema de programare liniară are cel puțin două soluții X_1 și X_2. În acest caz X_1 și X_2 trebuie să verifice relațiile (XXIV,3.3) și (XXIV,3.4), deci putem să scriem

$$AX_1 = b, \quad AX_2 = b,$$

$$X_1, \quad X_2 \geqslant 0.$$

Fie

$$X = \alpha X_1 + (1 - \alpha) X_2$$

o combinație convexă oarecare de X_1 și $X_2 (0 \leqslant \alpha \leqslant 1)$. Se vede că X este o soluție posibilă. Într-adevăr, avem

$$AX = A[\alpha X_1 + (1 - \alpha)X_2] = \alpha AX_1 + AX_2 - \alpha AX_2$$

sau

$$AX = \alpha b + b - \alpha b = b.$$

Pe de altă parte, se vede din expresia sa, că vectorul X este nenegativ.

În mod analog, teorema se demonstrează pentru cazul unui număr $n > 2$ de soluții posibile ale problemei de programare liniară.

Mulțimea punctelor de pe frontiera spațiului soluțiilor unei probleme de programare liniară este formată, după cum se poate constata ușor, din porțiuni ale hiperplanelor $AX = b$, presupunînd bineînțeles că spațiul soluțiilor problemei nu este o mulțime vidă. Prin urmare, spațiul soluțiilor unei probleme de programare liniară este o regiune din spațiul euclidian R_n, care poate fi sau vidă, sau o regiune convexă mărginită, sau o regiune convexă care este nemărginită într-o oarecare direcție. Dacă spațiul soluțiilor este o mulțime vidă, atunci este evident că, problema nu are soluții posibile și deci nu se poate vorbi de o soluție optimă; dacă spațiul soluțiilor este o regiune convexă mărginită, problema are soluții posibile și printre acestea se poate găsi o soluție optimă finită; dacă, însă, spațiul soluțiilor este o regiune convexă care este nemărginită într-o oarecare direcție, problema are soluții posibile, dar soluția optimă poate fi nemărginită. În cazul unei regiuni convexe mărginite (de exemplu, un poligon convex, un poliedru convex etc.) avem un număr finit de puncte extreme și, prin urmare, fiecare soluție posibilă este o combinație convexă de aceste puncte extreme. În cazul celălalt al unei regiuni convexe nemărginite avem de asemenea un număr finit de puncte extreme, însă, aici nu orice soluție posibilă poate fi exprimată printr-o combinație convexă de punctele extreme ale regiunii. Pentru simplificarea calculelor, vom considera numai problemele de programare liniară pentru care spațiul soluțiilor este o regiune convexă mărginită.

Prin demonstrația teoremei I, rezultă că, dacă o problemă de programare liniară are mai mult decît o singură soluție posibilă, problema are, în realitate, o infinitate de soluții posibile. Dintre mulțimea acestor soluții, urmează să găsim deci soluția optimă a problemei. Pentru aceasta ne bazăm pe teorema de mai jos.

TEOREMA II. *O soluție optimă a unei probleme de programare liniară corespunde la unul din punctele extreme ale spațiului soluțiilor. Dacă însă, soluția optimă corespunde la mai multe puncte extreme, ea are aceeași valoare pentru orice combinație convexă a acestor puncte.*

Vom demonstra teorema pentru cazul cînd spațiul soluțiilor problemei este un poligon convex din spațiul R_2. Demonstrația nu diferă esențial în cazul general. De asemenea, presupunem că soluția optimă este acea soluție posibilă care face minimă funcția obiectiv $F(X)$. Pentru cazul soluției optime care face maximă funcția obiectiv, demonstrația se păstrează cu schimbările corespunzătoare.

Scanned with OKEN Scanner

Fie, deci, $\overline{X}_1, \overline{X}_2, \ldots, \overline{X}_k$ vîrfurile poligonului convex, adică punctele extreme ale spaţiului soluţiilor problemei şi X_0 soluţia posibilă a problemei pentru care avem :

$$F(X) \geqslant F(X_0),$$

oricare ar fi X aparţinînd mulţimii soluţiilor posibile. Prima parte a teoremei este evidentă, dacă X_0 este el însuşi un punct extrem. Dacă însă X_0 nu este un punct extrem, atunci el se poate exprima printr-o combinaţie convexă a celor k puncte extreme ale poligonului convex, adică :

$$X_0 = \alpha_1\overline{X}_1 + \alpha_2\overline{X}_2 + \ldots + \alpha_k\overline{X}_k,$$

$$(\alpha_i \geqslant 0, \sum_{i=1}^{k} \alpha_i = 1),$$

de unde rezultă că

$$F(X_0) = F(\alpha_1\overline{X}_1 + \alpha_2\overline{X}_2 + \ldots + \alpha_k\overline{X}_k) = \alpha_1 F(\overline{X}_1) + \alpha_2 F(\overline{X}_2) + \ldots +$$

$$+ \alpha_k F(\overline{X}_k) = \min.$$

deoarece $F(X)$ este o funcţională liniară.

Presupunînd că minimul valorilor $F(\overline{X}_i)$, $(i = 1, 2, \ldots, k)$, este $F(\overline{X}_n)$ şi ţinînd seama că $\alpha_i \geqslant 0$ şi $\sum_{i=1}^{k} \alpha_i = 1$ putem să scriem :

$$F(X_0) \geqslant \alpha_1 F(\overline{X}_n) + \alpha_2 F(\overline{X}_n) + \ldots + \alpha_k F(\overline{X}_n)$$

sau

$$F(X_0) \geqslant F(\overline{X}_n).$$

Pe de altă parte, X_0 fiind soluţia pentru care avem $F(X) \geqslant F(X_0)$, pentru orice X aparţinînd spaţiului soluţiilor, deducem

$$F(X_0) = F(X_n) = \min$$

şi, prin urmare, punctul X_0 este un punct extrem al poligonului convex.

Rămîne acum să demonstrăm partea a doua a teoremei. Pentru aceasta presupunem că minimul căutat corespunde nu numai unui punct extrem, ci la un număr de g puncte extreme $\overline{X}_1, \overline{X}_2, \ldots, \overline{X}_g$, $(g \leqslant k)$. Cu alte cuvinte avem egalităţile :

$$F(\overline{X}_1) = F(\overline{X}_2) = \ldots = F(\overline{X}_g) = \min.$$

Dacă

$$X = \alpha_1\overline{X}_1 + \alpha_2\overline{X}_2 + \ldots \alpha_g\overline{X}_g, (\alpha_i \geqslant 0, \sum_{i=1}^{g} \alpha_i = 1),$$

este o combinaţie convexă oarecare de punctele extreme considerate, avem

$$F(X) = F(\alpha_1\overline{X}_1 + \alpha_2\overline{X}_2 + \ldots + \alpha_g\overline{X}_g) = \alpha_1 F(\overline{X}_1) + \alpha_2 F(\overline{X}_2) + \ldots \alpha_g F(\overline{X}_g)$$

Scanned with OKEN Scanner

sau

$$F(X) = \min,$$

şi, deci, teorema este demonstrată în întregime.

Potrivit acestei teoreme, pentru aflarea soluţiei optime a unei probleme de programare liniară, urmează să considerăm astfel numai punctele extreme ale mulţimii convexe a soluţiilor posibile ale problemei.

Se justifică astfel pentru care motiv, am determinat soluţia optimă, prin metoda grafică, calculînd valorile funcţiei obiectiv în punctele extreme. Este evident însă că această metodă are o aplicaţie mai restrînsă, neputînd fi practic aplicată cînd numărul variabilelor este mare.

Vom da acum cîteva indicaţii matematice, care explică procedeul întrebuinţat la metoda Simplex. Să presupunem pentru simplificare patru variabile x_1, x_2, x_3, x_4.

Ni se cere să facem maximă funcţia obiectiv

$$\Phi = t_1x_1 + t_2x_2 + t_3x_3 + t_4x_4 \qquad \text{(XXIV, 3.5)}$$

ştiind că

$$\begin{aligned} &a_{11}x_1 + a_{12}x_2 + a_{13}x_3 + a_{14}x_4 = a_1,\\ &a_{21}x_1 + a_{22}x_2 + a_{23}x_3 + a_{24}x_4 = a_2,\\ &a_{31}x_1 + a_{32}x_2 + a_{33}x_3 + a_{34}x_4 = a_3,\\ &x_1 \geqslant 0,\ x_2 \geqslant 0,\ x_3 \geqslant 0,\ x_4 \geqslant 0, \end{aligned} \qquad \text{(XXIV, 3.6)}$$

Evident, relaţiile de mai sus pot fi scrise sub o formă mai strînsă. Notăm:

$$V_0 = \begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix}, \quad V_1 = \begin{pmatrix} a_{11} \\ a_{21} \\ a_{31} \end{pmatrix}, \quad V_2 = \begin{pmatrix} a_{12} \\ a_{22} \\ a_{32} \end{pmatrix}, \quad V_3 = \begin{pmatrix} a_{13} \\ a_{23} \\ a_{33} \end{pmatrix}, \quad V_4 = \begin{pmatrix} a_{14} \\ a_{24} \\ a_{34} \end{pmatrix}.$$

Relaţiile (XXIV, 3.6) pot fi scrise:

$$V_0 = \sum_{i=1}^{4} x_i V_i. \qquad \text{(XXIV, 3.7)}$$

Întocmim tabelul 125, de la care pornim în aplicarea metodei Simplex.

Tabelul 125

Z / S	Baza	—	t_1	t_2	t_3	t_4	0	0	0
		V_0	V_1	V_2	V_3	V_4	V_5	V_6	V_7
0	V_5	a_1	a_{11}	a_{12}	a_{13}	a_{14}	1	0	0
0	V_6	a_2	a_{21}	a_{22}	a_{23}	a_{24}	0	1	0
0	V_7	a_3	a_{31}	a_{32}	a_{33}	a_{34}	0	0	1
S_i		0	0	0	0	0	0	0	1
$S_i - t_i$			$-t_1$	$-t_2$	$-t_3$	$-t_4$	0	0	0

Scanned with OKEN Scanner

Vectorii $V_i (i = 1, 2, 3, 4)$ nu sînt liniar independenți. Se știe că în spațiul liniar format din vectorii V_i, cu trei dimensiuni, dacă se ia o bază formată din trei vectori liniari independenți, toți ceilalți vectori ai spațiului se exprimă liniar în raport cu vectorii de bază.

Notînd :

$$V_5 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \quad V_6 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \quad V_7 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$$

putem scrie

$$\begin{aligned} V_0 &= a_1 V_5 + a_2 V_6 + a_3 V_7, \\ V_1 &= a_{11} V_5 + a_{21} V_6 + a_{31} V_7, \\ V_2 &= a_{12} V_5 + a_{22} V_6 + a_{32} V_7, \\ V_3 &= a_{13} V_5 + a_{23} V_6 + a_{33} V_7, \\ V_4 &= a_{14} V_5 + a_{24} V_6 + a_{34} V_7. \end{aligned} \qquad \text{(XXIV, 3.8)}$$

Numerele (a_1, a_2, a_3) sînt coordonatele vectorului V_0, în baza V_5, V_6, V_7, numerele (a_{11}, a_{21}, a_{31}), sînt coordonatele vectorului V_1 în baza V_5, V_6, V_7 și așa mai departe.

În tabel am așezat vectorii V_0, V_1, V_2, V_3, V_4, cu coordonatele lor în baza V_5, V_6, V_7. În prima linie a tabelului am trecut coeficienții t_i, $(i = 1,2, \ldots, 7)$, care intră în funcția obiectiv.

În prima coloană am așezat valorile t_i, corespunzătoare vectorilor bază V_5, V_6, V_7. Aceste valori sînt evident zero. În penultima linie așezăm valorile S_i, obținute înmulțind elementele din prima coloană cu elementele din coloana vectorului V_i. Evident, am obținut valorile 0. În ultima linie sînt așezate valorile $S_i - t_i (i = 1, 2, \ldots, 7)$. Toate aceste valori sînt negative. Să presupunem că de exemplu t_3 este cel mai mic dintre numerele t_i, diferite de zero. Vom introduce vectorul corespunzător V_3 în bază. Rămîne de văzut în locul căruia din vectorii de bază V_5, V_6, V_7 vom introduce vectorul V_3.

Dacă V_3 vine în locul lui V_5, din relațiile (XXIV, 3.8) deducem

$$V_5 = \frac{1}{a_{13}} V_3 - \frac{a_{23}}{a_{13}} V_6 - \frac{a_{33}}{a_{13}} V_7$$

și înlocuind această valoare în expresia corespunzătoare a lui V_0 din (XXIV, 3.8) rezultă :

$$V_0 = \frac{a_1}{a_{13}} V_3 + \left(a_2 - a_1 \frac{a_{23}}{a_{13}}\right) V_6 + \left(a_3 - a_1 \frac{a_{33}}{a_{13}}\right) V_7. \qquad \text{(XXIV, 3.9)}$$

Prin urmare, dacă în loc de baza V_5, V_6, V_7 punem baza (V_3, V_6, V_7), obținem pentru V_0 dezvoltarea (XXIV, 3.9).

Vom prefera să scriem pe V_0 sub forma următoare :

$$V_0 = \frac{a_1}{a_{13}} V_3 + a_{23}\left(\frac{a_2}{a_{23}} - \frac{a_1}{a_{13}}\right) V_6 + a_{33}\left(\frac{a_3}{a_{33}} - \frac{a_1}{a_{13}}\right) V_7. \qquad \text{(XXIV, 3.10)}$$

Scanned with OKEN Scanner

Dacă înlocuim baza V_5, V_6, V_7 cu V_5, V_3, V_7, obținem în același mod

$$V_0 = a_{13}\left(\frac{a_1}{a_{13}} - \frac{a_2}{a_{23}}\right)V_5 + \frac{a_2}{a_{23}}V_3 + a_{33}\left(\frac{a_3}{a_{33}} - \frac{a_2}{a_{23}}\right)V_7. \quad \text{(XXIV, 3.11)}$$

În cazul cînd în locul bazei V_5, V_6, V_7 punem baza V_5, V_6, V_3, rezultă

$$V_0 = a_{13}\left(\frac{a_1}{a_{13}} - \frac{a_3}{a_{33}}\right)V_5 + a_{23}\left(\frac{a_2}{a_{23}} - \frac{a_3}{a_{33}}\right)V_6 + \frac{a_3}{a_{33}}V_3, \quad \text{(XXIV, 3.12)}$$

V_0 trebuie să aibă coordonatele nenegative, deoarece aceste coordonate ne dau o soluție posibilă a sistemului considerat.

Să considerăm numerele $\frac{a_1}{a_{13}}$, $\frac{a_2}{a_{23}}$, $\frac{a_3}{a_{33}}$. Fie de exemplu, $\frac{a_3}{a_{33}}$, cel mai mic dintre ele. Rezultă atunci că V_0 scris sub forma (XXIV, 3.12) nu are coordonate negative.

Celelalte două formule (XXIV,3.10) și (XXIV,3.11), dacă $\frac{a_3}{a_{33}}$ este mai mic decît $\frac{a_1}{a_{13}}$ și $\frac{a_2}{a_{23}}$, conduc la coordonate inadmisibile pentru V_0, deoarece cel puțin una din coordonate este negativă. În cazul de față, vom schimba deci baza V_5, V_6 V_7 prin V_5, V_6, V_3.

Am demonstrat că putem introduce întotdeauna un vector în bază, în locul altuia, fără ca V_0 să capete coordonate negative.

Prin introducerea bazei V_5, V_6, V_3 vom aplica pentru V_0 formula (XXIV, 3,12) pentru care funcția obiectiv ia valoarea $\frac{a_3}{a_{33}}t_3 > 0$ deci a crescut față de valoarea zero, cît avea la început.

Obținem acum al doilea tabel din metoda Simplex (tabelul 126).

Tabelul 126

S \ Z	Baza	—	t_1	t_2	t_3	t_4	0	0	0
		V_0	V_1	V_2	V_3	V_4	V_5	V_6	V_7
0	V_5	b_1	b_{11}	b_{12}	0	b_{13}	1	0	b_{14}
0	V_6	b_2	b_{21}	b_{22}	0	b_{23}	0	1	b_{24}
t_3	V_3	b_3	b_{31}	b_{32}	1	b_{33}	0	0	b_{34}
S_i		b_3t_3	$b_{31}t_3$	$b_{32}t_3$	t_3	$b_{33}t_3$	0	0	$b_{34}t_3$

Numerele din acest tabel au fost obținute din tabelul 125 trecînd de la baza V_5, V_6, V_7, la baza V_5, V_6, V_3 pe care am văzut cum am ales-o. Ele se pot obține folosind relațiile (XXIV,3.8); o cale mai simplă se obține în baza observațiilor următoare:

a) înmulțirea tuturor elementelor din aceeași linie din tabelul 125 cu același număr revine la o schimbare de bază a vectorilor V_i.

Scanned with OKEN Scanner

b) scăderea din elementele unei linii a elementelor respective din altă linie înmulțite cu același număr revine la o schimbare de bază a vectorilor V_i.

Prin urmare, vom putea aplica regulile *a*) și *b*) astfel ca vectorul V_3 să ajungă să aibă coordonatele (0,0,1), iar V_5 și V_6 să rămînă cu coordonatele respective (1,0,0), (0,1,0). Pentru aceasta va fi suficient să împărțim mai întîi toate elementele din linia V_7 a tabloului 125 cu a_{33}. Linia astfel obținută, înmulțită respectiv cu a_{13} și a_{23}, o scădem din V_5 și V_6.

Adică elementele noului tabel se obțin din formulele:

$$b_{3j} = \frac{a_{3j}}{a_{33}},$$

$$b_{1j} = a_{1j} - \frac{a_{3j}}{a_{33}} a_{13}, \qquad 1 \leqslant j \leqslant 4,$$

$$b_{2j} = a_{2j} - \frac{a_{3j}}{a_{33}} a_{23}.$$

Pentru coordonatele vectorului V_0, am avut în mod analog:

$$b_3 = \frac{a_3}{a_{33}}$$

$$b_1 = a_1 - \frac{a_3}{a_{33}} a_{13}$$

$$b_2 = a_2 - \frac{a_3}{a_{33}} a_{23},$$

ceea ce ne permite să scriem dintr-o dată dacă notăm

$$b_{i0} = b_i \text{ și } a_{i0} = a_i,$$

$$b_{3j} = \frac{a_{3j}}{a_{33}},$$

$$b_{ij} = a_{ij} - \frac{a_{3j}}{a_{33}} a_{i3}, \qquad i \neq 3.$$

În general, dacă notăm V_l vectorul care va fi înlocuit, și cu V_k cel cu care vom completa noua bază, avem:

$$b_{kj} = \frac{a_{lj}}{a_{lk}},$$

$$b_{ij} = a_{ij} - \frac{a_{lj}}{a_{lk}} a_{ik}, \qquad i \neq l.$$

În tabelul 126 examinăm mărimea valorilor $S_i - t_i$. Cea mai mică dintre ele ne arată noul vector pe care trebuie să-l introducem în bază. El ia locul unui vector din bază astfel că coordonatele vectorului V_0 să nu devină negative. Repetăm, prin urmare, același procedeu și găsim un alt tabel.

Procedeul îl vom repeta pînă ce toate diferențele $S_i - t_i$ sînt pozitive. În acest caz, tabelul ne dă soluția optimă.

Să dovedim acest lucru.

Să presupunem că alegînd ca bază vectorul V_h, V_k, V_e, am obținut un tabel în care toate numerele $S_i - t_i$, sînt pozitive.

Fie (a_h, a_k, a_e), coordonatele vectorului V_0 în această bază:

$$V_0 = a_h V_h + a_k V_k + a_e V_e. \qquad \text{(XXIV, 3.13)}$$

Știm că $a_h \geqslant 0$, $a_k \geqslant 0$, $a_e \geqslant 0$.

De la baza V_h, V_k, V_e trecem la o altă bază V_m, V_n, V_p.

Fie respectiv (d_{mh}, d_{mk}, d_{me}), (d_{nh}, d_{nk}, d_{ne}), (d_{ph}, d_{pk}, d_{pe}), coordonatele vectorilor V_m, V_n, V_p.

Avem deci:

$$\begin{aligned} V_m &= d_{mh}V_h + d_{mk}V_k + d_{me}V_e, \\ V_n &= d_{nh}V_h + d_{nk}V_k + d_{ne}V_e, \\ V_p &= d_{ph}V_h + d_{pk}V_k + d_{pe}V_e. \end{aligned} \qquad \text{(XXIV, 3.14)}$$

Deoarece rezolvarea sistemului (XXIV, 3.14) este întotdeauna posibilă, deducem:

$$\begin{aligned} V_h &= D_{hm}V_m + D_{hn}V_n + D_{hp}V_p, \\ V_k &= D_{km}V_m + D_{kn}V_n + D_{kp}V_p, \\ V_e &= D_{em}V_m + D_{en}V_n + D_{ep}V_p, \end{aligned} \qquad \text{(XXIV, 3.15)}$$

obținînd de data aceasta coordonatele vectorilor

$$V_h, V_k, V_e, \text{ în baza } V_m, V_n, V_p.$$

Vectorul V_0, ținînd seama de (XXIV,3.13) devine

$$\begin{aligned} V_0 = {} & (a_h D_{hm} + a_k D_{km} + a_e D_{em})\, V_m + (a_h D_{hn} + a_k D_{kn} + a_e D_{en})\, V_n + \\ & + (a_h D_{hp} + a_k D_{kp} + a_e D_{ep})\, V_p. \end{aligned} \qquad \text{(XXIV, 3.16)}$$

Pe de altă parte, conform definiției avem

$$\begin{aligned} S_m &= d_{mh}t_h + d_{mk}t_k + d_{me}t_e, \\ S_n &= d_{nh}t_h + d_{nk}t_k + d_{ne}t_e, \\ S_p &= d_{ph}t_h + d_{pk}t_k + d_{pe}t_e, \end{aligned} \qquad \text{(XXIV, 3.17)}$$

din care rezultă

$$\begin{aligned} t_h &= D_{hm}S_m + D_{hn}S_n + D_{hp}S_p, \\ t_k &= D_{km}S_m + D_{kn}S_n + D_{kp}S_p, \\ t_e &= D_{em}S_m + D_{en}S_n + D_{ep}S_p. \end{aligned} \qquad \text{(XXIV, 3.18)}$$

Scanned with OKEN Scanner

Valoarea funcției obiectiv în baza (V_h, V_k, V_e) este

$$a_h t_h + a_k t_k + a_e t_e = (a_h D_{hm} + a_k D_{km} + a_e D_{em}) S_m +$$
$$+ (a_h D_{hn} + a_k D_{kn} + a_e D_{en}) S_n + (a_h D_{hp} + a_k D_{kp} + a_e D_{ep}) t_e$$

ținînd seama de relațiile precedente.

Valoarea funcției obiectiv, în baza (V_m, V_n, V_p), ținînd seama de relația (XXIV, 3.16), este:

$$(a_h D_{hm} + a_k D_{km} + a_e D_{em}) t_m + (a_h D_{hn} + a_k D_{kn} + a_e D_{en}) t_n +$$
$$+ (a_h D_{hp} + a_k D_{kp} + a_e D_{ep}) t_e .$$

Creșterea funcției obiectiv de la baza (V_h, V_k, V_e), la baza (V_m, V_n, V_p), este dată de relația:

$$(a_h D_{hm} + a_k D_{km} + a_e D_{em}) (t_m - S_m) +$$
$$+ (a_h D_{hn} + a_k D_{kn} + a_e D_{en}) (t_n - S_n) + \qquad \text{(XXIV, 3.19)}$$
$$+ (a_h D_{hp} + a_k D_{kp} + a_e D_{ep}) (t_p - S_p).$$

Coordonatele vectorului V_0 rezultînd din relația (XXIV,3.16) sînt întotdeauna pozitive sau nule. Valorile $t_m - S_m$, $t_n - S_n$, $t_p - S_p$ sînt negative sau nule. Diferența (XXIV, 3.19) este deci întotdeauna negativă sau nulă, ceea ce arată, ținînd seama că nu am impus nici o restricție vectorilor V_m, V_n, V_p, că maximul funcției obiectiv se obține în baza (V_h, V_n, V_e).

Metoda Onicescu. Prof. O. Onicescu a indicat o metodă de rezolvare numerică a sistemelor de ecuații liniare. Această metodă poate fi aplicată la rezolvarea problemelor de programare liniară.

M. Iosifescu, R. Theodorescu, L. Todor *), au arătat modul de aplicare al acestei metode, analizînd condițiile în care ea se aplică mai avantajos.

Se pleacă de la sistemul

$$\sum_{j=1}^{n} a_j^i x^j = b^i, \qquad 1 \leqslant i \leqslant n \qquad \text{(XXIV, 3.20)}$$

n care determinantul $(a_j^i) \neq 0$.

Să efectuăm o transformare nesingulară

$$\alpha_e^h = \sum_{j=1}^{n} a_e^j \lambda_j^h, \quad \beta^h = \sum_{j=1}^{n} b^j \lambda_j^h, \quad 1 \leqslant h, e \leqslant n, \qquad \text{(XXIV, 3.21)}$$

astfel încît sistemul

$$\sum_{j=1}^{n} \alpha_j^i x^j = \beta^i, \qquad 1 \leqslant i \leqslant n \qquad \text{(XXIV, 3.22)}$$

*) M. Iosifescu, R. Theodorescu, L. Todor, *Metoda Onicescu de reducere a sistemelor de ecuații liniare.* Comunicările Academiei R.P.R., **XI**, 5, 513 (1961).

Scanned with OKEN Scanner

să aibe proprietatea

$$\sum_{j=p+1}^{n} \alpha_j^k x^j = \xi \sum_{j=p+1}^{n} \alpha_j^k, \; 1 \leqslant k \leqslant p+1. \qquad \text{(XXIV, 3.23)}$$

Folosind această metodă soluția problemelor de programare liniară poate căpăta unele simplificări.

Enunțăm problema generală a programării liniare sub forma următoare: să se găsească maximul formei liniare

$$c_1x_1 + c_2x_2 + \ldots + c_nx_n,$$

în condițiile

$$\begin{aligned} a_{11}x_1 + \ldots + a_{1n}x_n &= b_1, \\ a_{21}x_1 + \ldots + a_{2n}x_n &= b_2, \qquad \text{(XXIV, 3.24)} \\ \ldots\ldots\ldots\ldots\ldots\ldots \\ a_{m1}x_1 + \ldots + a_{mn}x_n &= b_m, \qquad m < n \\ x_1 \geqslant 0 \ldots x_n &\geqslant 0. \end{aligned}$$

Admitem că matricea sistemului de mai sus are rangul m și că $x_1, x_2, \ldots, x_m$ sînt necunoscute principale.

Atunci relațiile de mai sus pot fi scrise:

$$\begin{aligned} a_{11}x_1 + \ldots + a_{1m}x_m &= b_1 - a_{1,m+1}x_{m+1} \ldots a_{1n}x_n, \\ a_{21}x_1 + \ldots + a_{2m}x_m &= b_2 - a_{2,m+1}x_{m+1} \ldots a_{2n}x_n, \qquad \text{(XXIV, 3.25)} \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ a_{m1}x_1 + \ldots + a_{mm}x_m &= b_m - a_{m,m+1}x_{m+1} \ldots a_{mn}x_n. \end{aligned}$$

Să luăm

$$\xi = c_{p+1}x_{p+1} + \ldots + c_mx_m$$

presupunînd că numerele c_i, $(i = p+1, \ldots, m)$, sînt pozitive: $(1 < p < m)$.

Metoda mai sus arătată este aplicabilă sistemului (XXIV,3.25).

În consecință, problema programării liniare se enunță astfel: să se găsească maximul formei liniare

$$c_1x_1 + \ldots + c_px_p + \xi + c_{m+1}x_{m+1} + \ldots + c_nx_n,$$

în condițiile

$$\begin{aligned} \sum_{j=1}^{p} \alpha_{1j}x_j + \alpha_1\xi &= \beta_1 - \sum_{j=m+1}^{n} \alpha_{1j}x_j, \\ \ldots\ldots\ldots\ldots\ldots\ldots \\ \sum_{j=i}^{p} \alpha_{p+1,j}x_j + \alpha_{p+1}\xi &= \beta_{p+1} \sum_{j=m+1}^{n} \alpha_{p+1,j}x_j, \qquad \text{(XXIV, 3.26)} \\ x_1 \geqslant 0_1 \ldots x_p \geqslant 0,\; \xi \geqslant 0,\; x_{m+1} &\geqslant 0 \ldots x_n \geqslant 0. \end{aligned}$$

Scanned with OKEN Scanner

Noua problemă de programare liniară are $n - m + p + 1$ necunoscute, deci mai puțin decît n necunoscute cîte are problema de la care am plecat.

Autorii citați au demonstrat că dacă problema inițială are o soluție, atunci și problema transformată are o soluție. Dacă cele două probleme admit o soluție, cele două soluții coincid numai dacă valorile $x_{p+1}, \ldots, x_m$, calculate din sistemul (XXIV, 3.26), după rezolvarea problemei transformate de programare liniară, sînt pozitive.

Dăm următorul exemplu, considerat în lucrarea citată.

Să se găsească maximul formei liniare.

$$L = 2x_1 + x_2 + 3x_3 - x_4 + 2x_5,$$

in condițiile

$$\begin{aligned} &x_1 + 2x_2 + x_3 + x_5 + x_6 = 10,\\ &x_2 + x_3 + x_4 + x_5 + x_7 = 8,\\ &x_1 + x_3 + x_4 + x_8 = 5,\\ &x_i \geqslant 0, \quad 1 \leqslant i \leqslant 8. \end{aligned}$$

Luăm

$$\xi = 2x_1 + 3x_3 + 2x_5 .$$

Prin urmare se pune problema ca funcția

$$L = \xi + x_2 - x_4$$

să fie maximă.

Adunînd ecuațiile de condiție, găsim,

$$\xi = 23 - 3x_2 - 2x_4 - x_6 - x_7 - x_8 .$$

Prin urmare, funcția obiectiv devine

$$L = 23 - 2x_2 - 3x_4 - x_6 - x_7 - x_8,$$

ceea ce ne conduce la soluția imediată

$$x_2 = 0,\ x_4 = 0,\ x_60,\ x_7 = 0,\ x_3 = 0.$$

Rezultă apoi

$$x_1 = 2,\ x_3 = 3,\ x_5 = 5,$$

adică soluția problemei inițiale de programare liniară.

4. Problema transporturilor

Să considerăm m centre de producție ale unui același produs :

$$A_1, A_2, \ldots, A_m .$$

Produsul respectiv trebuie să fie distribuit la n centre de consum

$$B_1, B_2, \ldots, B_n .$$

Scanned with OKEN Scanner

Notăm prin a_i cantitatea produsă în centrul A_i, $(i = 1, 2, \ldots, m)$, și prin b_j cantitatea care trebuie distribuită la centrul de consum B_j, $(j = 1, 2, \ldots, n)$.

Să admitem că pe baza unei statistici se stabilește că cheltuielile de transport pentru o unitate de produs (sau, eventual, timpul necesar pentru transportul unei tone-kilometru), sau pentru distanța de la A_i la B_j, reprezintă C_{ij} lei (ore, tone-km).

Dacă însemnăm prin X_{ij} cantitatea de produs transportată de la A_i la B_j, problema care ne-o punem este aceea de a alege astfel numerele X_{ij}, *încît costul transporturilor* (Z) *de la toate centrele* A_i *la* B_j *să fie cît mai mic.*

Ținînd seama de notațiile de mai sus, costul total al transporturilor (Z) este

$$Z = \sum_{i=1}^{m} \sum_{j=1}^{n} C_{ij} X_{ij}.$$

Prin urmare, trebuie să determinăm numerele X_{ij} astfel încît să avem

$$Z = \sum_{i=1}^{m} \sum_{j=1}^{n} C_{ij} X_{ij} = \text{minim.} \qquad \text{(XXIV, 4.1)}$$

De altă parte, condițiile concrete ale problemei puse, ne arată că:

a) totalul cantității de produs expediată de la A_i, este egal cu a_i; trebuie, deci, să avem

$$\sum_{j=1}^{n} X_{ij} = a_i, \quad (i = 1, 2 \ldots m); \text{ (XXIV, 4.2)}$$

b) suma cantităților primite la centrul B_j trebuie să fie b_j; prin urmare trebuie să mai avem

$$\sum_{i=1}^{m} X_{ij} = b_j, \quad (j = 1, 2, \ldots, n); \text{ (XXIV, 4.3)}$$

c) este necesar ca toate cantitățile X_{ij} să fie pozitive sau nule; rezultă deci și o a treia condiție

$$X_{ij} \geqslant 0. \qquad \text{(XXIV, 4.4)}$$

Am ajuns astfel la o formulare matematică a problemei cu caracter economic concret, practic, pe care o avem de rezolvat: *expresia liniară* $(XXIV, 4.1)$ *trebuie făcută minimă cu îndeplinirea simultană a condițiilor date de relațiile* $(XXIV, 4.2)$, $(XXIV, 4.3)$ *și* $(XXIV, 4.4)$.

Elementele problemei tip de programare liniară formulată mai sus pot fi ilustrate printr-un tabel-matrice ca în tabelul 127.

Scanned with OKEN Scanner

Tabelul 127

Centru de consum / Centru de producție	B_1	B_2	B_3	...	B_j	...	B_n	Producție
A_1	X_{11} C_{11}	X_{12} C_{12}	X_{13} C_{13}	...	X_{1j} C_{1j}	...	X_{1n} C_{1n}	a_1
A_2	X_{21} C_{21}	X_{22} C_{22}	X_{23} C_{23}	...	X_{2j} C_{2j}	...	X_{2n} C_{2n}	a_2
A_3	X_{31} C_{31}	X_{32} C_{32}	X_{33} C_{33}	...	X_{3j} C_{3j}	...	X_{3n} C_{3n}	a_3
⋮					⋮			
A_i	X_{i1} C_{i1}	X_{i2} C_{i2}	X_{i3} C_{i3}	...	X_{ij} C_{ij}	...	X_{in} C_{in}	a_i
⋮	⋮				⋮			
A_m	X_{m1} C_{m1}	X_{m2} C_{m2}	X_{m3} C_{m3}	...	X_{mj} C_{mj}	...	X_{mn} C_{mn}	a_m
Consum	b_1	b_2	b_3	...	b_j	...	b_n	Σ

Fiecare rubrică din acest tabel, cu excepția celor de pe prima și ultima linie, respectiv prima și ultima coloană, conțin cîte doi indicatori: unul — notat prin X_{ij} (în colțul din stînga de jos al fiecărei rubrici) — reprezintă cantitățile *necunoscute* de produs care trebuie transportate de la A_i la B_j, satisfăcînd condiția (XXIV,4.1) și condițiile suplimentare (XXIV,4.2), (XXIV,4.3) și (XXIV,4.4); iar al doilea — notat prin C_{ij} (în colțul din dreapta sus) — reprezintă *coeficientul constant* (cheltuiala pe unitatea de produs sau distanța de la centrul A_i la centrul B_j, în caz că minimul căutat ar fi numărul total de tone-kilometri transportați), cu ajutorul căruia se apreciază eficacitatea economică de utilizare a unei unități de produs în raport cu destinația ce i-a fost stabilită.

În ultima coloană sînt date cantitățile de producție ale centrelor A_i care trebuie transportate la centrele B_j, iar pe ultima linie, cantitățile de producție ce trebuie recepționate la centrele de consum de la diferite centre de producție.

După modelul acestui tabel teoretic vom da un exemplu de rezolvare a problemei transporturilor.

Pentru simplificarea calculelor, considerăm 3 centre de producție P_1, P_2, P_3 și 4 centre de consum C_1, C_2, C_3, C_4.

La centrele de producție avem următoarele cantități de materiale :

— centrul de producție P_1, are 70 000 tone
— „ „ „ P_2 „ 10 000 „
— „ „ „ P_3 „ 20 000 „
Total 100 000 tone

iar la centrele de consum este nevoie de același material în următoarele cantități :

— centrul de consum C_1 are nevoie de 50 000 tone
— „ „ „ C_2 „ „ „ 25 000 „
— „ „ „ C_3 „ „ „ 15 000 „
— „ „ „ C_4 „ „ „ 10 000 „

Tabelul cheltuielilor de transport al unei tone de la fiecare centru de producție la fiecare centru de consum se caracterizează prin următoarele date (tabelul 128).

Tabelul 128

Consumator / Producător	C_1	C_2	C_3	C_4	Resurse de materiale (mii tone)
P_1	3	2	2	4	70
P_2	1	2	3	4	10
P_3	3	5	2	1	20
Consum	50	25	15	10	100

Din acest tabel se vede, de pildă, că transportul unei tone de material de la centrul de producției P_3 la centrul de consum C_2 costă 5 unități bănești (lei, sute de lei, mii lei etc.); sau de la centrul de producție P_1 la centrul de consum C_4 costă 4 unități.

Ni se cere să elaborăm un plan de transport care să necesite cheltuieli minime, indiferent de la ce centru de producție, la ce centru de consum se efectuează transportul.

Dacă din centrul de producție i la centrul de consum j, transportăm cantitatea x_{ij}, atunci problema de programare liniară constă în a găsi minimul funcției liniare :

$$3x_{11} + 2x_{12} + 2x_{13} + 4x_{14} + x_{21} + 2x_{22} + 3x_{23} + \\ + 4x_{24} + 3x_{31} + 5x_{32} + 2x_{33} + x_{34} = \sum_{i=1}^{m} \sum_{j=1}^{n} Z_{ij} x_{ij} \qquad (XXIV, 4.5)$$

(notînd prin Z_{ij} — costul unei tone de transport de la centrul de producție i la centrul de consum j).

Este evident că minimul căutat al funcției trebuie găsit în condiția respectării simultane a condițiilor restrictive formulate în enunțul problemei :

$$\begin{aligned} x_{11} + x_{12} + x_{13} + x_{14} &= 70, \\ x_{21} + x_{22} + x_{23} + x_{24} &= 10, \\ x_{31} + x_{32} + x_{33} + x_{34} &= 20, \\ x_{11} + x_{21} + x_{31} &= 50, \\ x_{12} + x_{22} + x_{32} &= 25, \\ x_{13} + x_{23} + x_{33} &= 15, \\ x_{14} + x_{24} + x_{34} &= 10. \end{aligned} \qquad \text{(XXIV, 4.6)}$$

Pentru rezolvarea acestei probleme, folosirea procedeului obișnuit al metodei simplex ar duce la complicații inutile. Ea se poate rezolva mai ușor prin așa-numita *metodă a repartiției*.

Aplicînd această metodă vom ține seama de următoarele :

a) Dacă numărul centrelor de producție (expediere) este m iar numărul centrelor de consum n, atunci în planul optim ca și în planurile intermediare vor figura $m + n - 1$ termeni.

b) Tabelul inițial de calcul poate fi obținut prin *metoda diagonalei*. Aceasta înseamnă că el se completează începînd de la colțul din stînga de sus și terminînd cu colțul din dreapta de jos. Pînă la completare, un asemenea tabel are următoarea înfățișare (tabelul 129).

Tabelul 129

	C_1	C_2	C_3	C_4	
P_1	3	2	2	4	70
P_2	1	2	3	4	10
P_3	3	5	2	1	20
	50	25	15	10	100

În chenarele din colțurile din dreapta sus ale fiecărei rubrici — după cum se vede — figurează coeficienții constanți : (cheltuielile de transport), urmînd ca volumul mărfurilor transportate să fie completat. De pildă, cantitatea ce va fi transportată de la centrul P_1 la centrul C_1 se va înscrie în primul pătrat din colțul din stînga sus. Este clar că poate fi vorba de maximum 50 de mii tone, de cîte are nevoie centrul C_1. Dacă înscriem întreaga cantitate, atunci, desigur, prima coloană este exclusă din calculele următoare. În acest caz centrului de producție P_1 i-ar mai rămîne o can-

titate de 20 mii tone pe care o vom destina celui de-al doilea centru de consum C_2.

Un asemenea raţionament ne conduce „pas cu pas" la tabelul 130.

Tabelul 130

	C_1	C_2	C_3	C_4	
P_1	3 / 50	2 / 20	2	4	70
P_2	1	2 / 5	3 / 5	4 / 0	10
P_3	3	2	2 / 10	1 / 10	20
	50	25	15	10	100

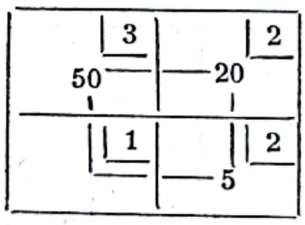
După cum se vede, în pătratul P_2, C_4 s-a pus cifra 0 care arată că nu avem ce transporta.

Tabelul 130 cuprinde un plan de transport iniţial.

Vom continua acum să-l îmbunătăţim.

c) Pentru a întocmi o nouă variantă îmbunătăţită, trebuie să examinăm pătratele rămase necompletate. Fiecărui pătrat liber îi corespunde un ciclu închis şi numai unul, format din linii orizontale şi verticale, unul din colţurile acestui ciclu este chiar în pătratul liber, restul colţurilor aflîndu-se în pătrate completate.

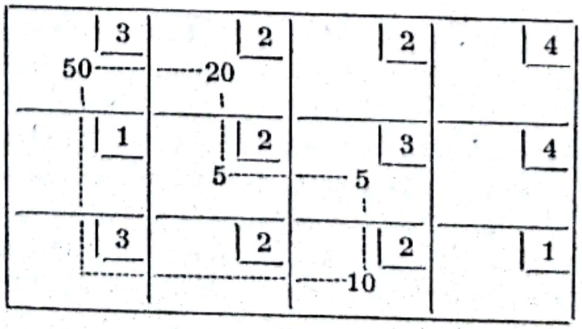
De pildă, la pătratul P_2, C_1 se referă următorul ciclu :



Observăm că acest ciclu are patru colţuri.

Pentru pătratul liber P_3, C_1, avem alt ciclu care are şase colţuri (tabelul 131).

Tabelul 131



Mai departe, alternînd semnele, se face suma algebrică, a coeficienților transportului — corespunzători fiecărui colț al ciclului format — ținînd seama ca să punem semnul plus coeficientului din colțul ciclului aflat în pătratul liber. De pildă, pentru primul ciclu format această sumă va fi:

$$1 - 2 + 2 - 3 = -2, \text{ iar pentru ciclul al doilea:}$$

$$3 - 2 + 3 - 2 + 2 - 3 = 1.$$

Efectuînd acest calcul pentru toate pătratele libere vom obține tabelul 132.

Tabelul 132

3 50	2 20	2 [−1]	4 [0]
1 [−2]	2 5	3 5	4 0
3 [1]	5 [4]	2 10	1 10

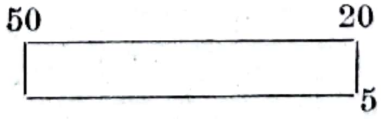
În acest tabel sumele algebrice referitoare la pătratele goale sînt plasate în chenare punctate. Locul liber în care suma algebrică este negativă poate fi inclus în program.

În exemplul nostru avem două asemenea valori negative. Îmbunătățirea planului se efectuează în modul următor:

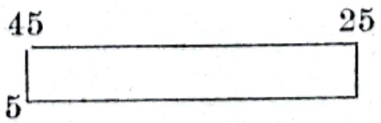
Mai întîi îmbunătățim planul în raport cu valoarea negativă din pătratul P_2, C_1. Procedăm în felul următor:

Din cantitățile de transportat care figurează în pătratele în care termenii sumei algebrice din ciclul respectiv au fost negativi (în cazul nostru din cantitățile de 50 000 tone și 5 000 tone), se ia cantitatea cea mai mică (5 000 tone) și se trece în pătratele în care termenii sumei algebrice au fost pozitivi și totodată se scade din pătratele cu colțurile ciclului negative.

În cazul nostru, ciclul:



se transformă în ciclul:



Acum s-a creat un alt loc liber și este posibil să se elaboreze un alt plan de transport ca în tabelul 133.

Tabelul 133

3 / 45	2 / 25	2	4
1 / 5	2	3 / 5	4 / 0
3	5	2 / 10	1 / 10

Pentru a verifica optimalitatea acestui plan vom proceda din nou la formarea ciclurilor pentru locurile libere și vom calcula sumele algebrice ale coeficienților după regula cunoscută pentru a vedea dacă mai găsim vreo valoare negativă care ne va da indicație că planul de transport mai poate fi îmbunătățit.

Făcînd calculele respective obținem următorul tabel de control (tabelul 134).

Tabelul 134

3 / 45	2 / 25	2 / [5]	4 / [−2]
1 / 5	2 / [2]	3 / 5	4 / 0
3 / [3]	5 / 0	2 / 10	1 / 10

Se vede că în pătratul P_1, C_4 avem un element negativ ceea ce înseamnă că acest pătrat mai poate fi folosit pentru îmbunătățirea planului.

Trecînd, din cantitățile aflate în pătratele cu colțurile ciclului negative (45 și 0), pe cea mai mică cantitate (0) în pătratele cu colțurile negative, scăzînd-o în același timp din cantitatea aflată în pătratul cu colț pozitiv, planul de transport capătă forma din tabelul 135.

Tabelul 135

3 / 45	2 / 25	2 / [−3]	4 / 0
1 / 5	2	3 / 5	4 / 0
3	5 / 0	2 / 10	1 / 10

Planul, după cum se vede, a rămas neschimbat, deoarece am scăzut şi am adăugat numărul 0.

Pentru pătratul P_1C_3, considerînd ciclul P_1C_3, P_1C_1, P_2C_1, P_2C_3, avem suma algebrică $2 - 3 + 1 - 3 = -3 < 0$. Îmbunătăţim planul scăzînd pe 5 din 45 şi 5, şi adăugînd această valoare la 5 şi 0, obţinem planul din tabelul 136.

Tabelul 136

40 | 3	25 | 2	5 | 2	0 | 4
10 | 1	0 | 2	0 | 3	0 | 4
0 | 3	0 | 5	10 | 2	10 | 1

Pentru acest plan nu mai găsim un ciclu care să ne conducă la valori negative. Deci nu mai putem să-l îmbunătăţim.

Cheltuielile de transport după acest program optim de transport sînt :

$$40 \cdot 3 + 25 \cdot 2 + 5 \cdot 2 + 10 \cdot 1 + 10 \cdot 2 + 10.1 = 220$$ unităţi băneşti (mii lei).

Faţă de primul plan care a necesitat un volum de cheltuieli de

$$50 \cdot 3 + 20 \cdot 2 + 5 \cdot 2 + 5 \cdot 3 + 10 \cdot 2 + 10 \cdot 1 = 245$$

unităţi băneşti, planul optim prezintă o economie de 25 unităţi băneşti. Organizarea unui alt plan de transport mai avantajos din punct de vedere economic în condiţiile date, nu este posibilă.

Metoda folosită la rezolvarea problemei este tocmai *metoda repartiţiei.*

Justificarea matematică a acestei metode este simplă. Să notăm prin C_{ij} preţul de transport pe unitate de la centrul de producţie P_i, $(i = 1,2, \ldots, m)$, la locul de consum C_j, $(j = 1,2, \ldots, n)$. Fie x_{ij} cantitatea transportată de la P_i la C_j. Trebuie ca funcţia

$$U = \sum_{i,j} C_{ij} x_{ij}$$

să fie făcută minimă ştiind că sumele

$$\sum x_{ij} = \alpha_i \quad \text{şi} \quad \sum_i x_{ij} = \beta_j \qquad \text{(XXIV,4.7)}$$

sînt constante.

Să considerăm ca exemplu un ciclu de şase valori pentru care avem :

$$C_{i_1 j_1} - C_{i_2 j_2} + C_{i_3 j_3} - C_{i_4 j_4} + C_{i_5 j_5} - C_{i_6 j_6} < 0. \qquad \text{(XXIV,4.8)}$$

Suma corespunzătoare din U pentru acest ciclu este :

$$U_1 = C_{i_2 j_2} x_{i_2 j_2} + C_{i_3 j_3} x_{i_3 j_3} + C_{i_4 j_4} x_{i_4 j_4} + C_{i_5 j_5} x_{i_5 j_5} + C_{i_6 j_6} x_{i_6 j_6} .$$

Scanned with OKEN Scanner

Să presupunem că dintre numerele $x_{i_2j_2}$, $x_{i_4j_4}$, $x_{i_6j_6}$, care corespund la cicluri cu valori negative, cel mai mic este $x_{i_6j_6}$. În loc de cantităţile transportate

$$x_{i_1j_1} = 0,\ x_{i_2j_2},\ x_{i_3j_3},\ x_{i_4j_4},\ x_{i_5j_5},\ x_{i_6j_6}$$

vom pune acum cantităţile

$$y_{i_1j_1},\ y_{i_2j_2},\ y_{i_3j_3},\ y_{i_4j_4},\ y_{i_5j_5},\ y_{i_6j_6}$$

după regula pe care am folosit-o pînă acum.

$$y_{i_1j_1} = x_{i_6j_6},\ y_{i_3j_3} = x_{i_3j_3} + x_{i_6j_6},\ y_{i_5j_5} = x_{i_5j_5} + x_{i_6j_6},$$
$$y_{i_2j_2} = x_{i_2j_2} - x_{i_6j_6},\ y_{i_4j_4} = x_{i_4j_4} - x_{i_6j_6},$$
$$y_{i_6j_6} = x_{i_6j_6} - x_{i_6j_6} = 0.$$

Suma U_1 va fi înlocuită prin suma U_2:

$$U_2 = C_{i_1j_1}\, y_{i_1j_1} + C_{i_2j_2}\, y_{i_2j_2} + \ldots + C_{i_6j_6}\, y_{i_6j_6}.$$

Se vede însă că

$$U_2 = U_1 + (C_{i_1j_1} - C_{i_2j_2} + C_{i_3j_3} - C_{i_4j_4} + C_{i_5j_5} - C_{i_6j_6})\, x_{i_6j_6}.$$

Ţinînd seama de relaţia (XXIV,3.4) se vede că

$$U_2 < U_1,$$

ceea ce înseamnă că înlocuind în U pe U_1 prin U_2, obţinem o îmbunătăţire a soluţiei.

Este esenţial de observat că prin acest procedeu suma cantităţilor transportate, atît pe linii, cît şi pe coloane, rămîne neschimbată. Într-adevăr, pe fiecare linie şi fiecare coloană avem în sumele algebrice ale ciclurilor un număr C_{ij} pozitiv şi altul negativ. La sumele x_{ij} corespunzătoare adăugăm aceeaşi cantitate şi o scădem, ceea ce înseamnă că totalul sumei transportate pe linii şi coloane rămîne acelaşi. Obţinem astfel îmbunătăţiri succesive, rămînînd în condiţiile problemei. O analiză, pe care nu o mai facem aici, ne arată că prin aceste ameliorări succesive, ajungem la programul optim.

Interesant de observat este că dacă

$$C_{i_1j_1} - C_{i_2j_2} + C_{i_3j_3} - C_{i_4j_4} - C_{i_5j_5} + C_{i_6j_6} = 0 \qquad \text{(XXIV, 4.9)}$$

se obţine, oricare ar fi $x_{i_6j_6}$, îndeplinind condiţiile (XXIV,4.7),

$$U_2 = U_1.$$

Prin urmare, dacă există o relaţie (XXIV,4.9) şi avem realizat un program optim, în baza raţionamentului rezultă că avem o infinitate de programe optime. Putem verifica acest lucru pe exemplul ales.

Reamintim că valorile C_{ij} sînt următorarele:

	C_1	C_2	C_3	C_4
P_1	3	2	2	4
P_2	1	2	3	4
P_3	3	5	2	1

Observăm că pentru pătratele P_1C_1, P_1C_3, P_3C_3, P_3C_1, avem ciclul

$$3 - 2 + 2 - 3 = 0.$$

Înseamnă, în baza observației de mai sus, că nu avem o singură soluție optimă, ci o infinitate de soluții.

Se poate arăta direct acest lucru pe o cale elementară. Aceasta datorită faptului că avem un număr mic de date.

Din relațiile (XXIV,4.6) deducem

$$\begin{aligned} x_{11} &= 45 - x_{13} - x_{21} - x_{23} + x_{32} + x_{34}, \\ x_{14} &= 25 - x_{12} + x_{21} + x_{23} - x_{32} - x_{34}, \\ x_{22} &= 25 - x_{12} - x_{32}, \\ x_{24} &= -15 + x_{12} - x_{21} - x_{23} + x_{32}, \\ x_{31} &= 5 + x_{13} + x_{23} - x_{32} - x_{34}, \\ x_{33} &= 15 - x_{13} - x_{23}. \end{aligned} \qquad \text{(XXIV, 4.10)}$$

Avem 7 relații (XXIV,4.6), dintre care ultima este consecința celorlalte. Prin urmare, între cele 12 valori x_{ij}, avem 6 relații liniare distincte. Urmează că 6 dintre ele și anume x_{12}, x_{13}, x_{21}, x_{23}, x_{32}, x_{34} pot fi socotite ca variabile independente, iar celelalte funcție de ele, date de relațiile (XXIV,4.10).

Funcția U devine în acest caz, ținînd seama de (XXIV,4.5),

$$U = 270 - 2x_{21} + x_{23} + 3x_{32} - 3x_{34}, \qquad \text{(XXIV, 4.11)}$$

unde x_{21}, x_{23}, x_{32}, x_{34} sînt variabile independente îndeplinind condițiile:

$$0 \leqslant x_{21} \leqslant 10,\ 0 \leqslant x_{23} \leqslant 10,\ 0 \leqslant x_{32} \leqslant 20,\ 0 \leqslant x_{34} \leqslant 10, \qquad \text{(XXIV, 4.12)}$$

decurgînd din (XXIV,4.6).

Pentru ca valoarea lui U, dată de (XXIV,4.11) să fie minimă trebuie ca x_{23} și x_{32} să ia cea mai mică valoare posibilă și x_{21} și x_{34} cea mai mare valoare posibilă.

Trebuie deci, ținînd seama de (XXIV,4.12), să avem:

$$x_{23} = 0, \quad x_{32} = 0, \quad x_{21} = 10, \quad x_{34} = 10.$$

Dacă $x_{21} = 10$, din a doua din egalitățile (XXIV,4.6) deducem $x_{22} = 0$, $x_{24} = 0$.

Din $x_{34} = 10$ și ultima din egalitățile (XXIV,4.6) deducem

$$x_{14} = 0, \quad x_{24} = 0.$$

Notînd $x_{13} = \lambda$, completăm în același mod tabloul valorilor x_{ij} și obținem:

	C_1	C_2	C_3	C_4
P_1	$45-\lambda$ | 3	25 | 2	λ | 2	0 | 4
P_2	10 | 1	0 | 2	0 | 3	0 | 4
P_3	$\lambda-5$ | 3	0 | 5	$15-\lambda$ | 2	10 | 1

(XXIV, 4.13)

λ fiind un parametru care variază între 5 și 15 inclusiv.

Tabloul (XXIV,4.13) ne dă programul optim. Cum cantitățile de transportat C_{ij} pot să nu fie numere întregi, rezultă din tabloul (XXIV, 4.13) că există o infinitate de programe optime.

Din cele de mai sus am putut constata că problema transporturilor reprezintă un caz special de programare liniară în care toți coeficienții ecuațiilor restrictive sînt egali cu 1 sau 0. Ea poate fi rezolvată prin metoda simplex însă calea de urmat este mai complicată și atunci s-au găsit algoritme speciale mai ușor de întrebuințat.

În afară de metoda repartiției sau „în scară" prezentată mai sus s-au mai elaborat în ultimul timp și alte metode interesante, însă, conținînd aceeași idee de rezolvare :

— în prima etapă se caută o primă soluție posibilă cît mai apropiată de soluția optimă

— în etapa a doua, soluția inițială este treptat îmbunătățită. Un interesant procedeu se aplică și la noi în țară la Ministerul Economiei Forestiere, începînd din luna iunie 1960 *).

El reprezintă — după cum arată autorii lui — o combinație a două metode, dintre care prima — așa-zisa metodă a „diferenței maxime" servește la găsirea unei soluții inițiale, iar cea de-a doua — este o variantă a metodei „în scară" care servește la îmbunătățirea soluției inițiale prin iterații succesive. Rezultă, deci, că potrivit acestui procedeu, în prima etapă se determină anumite valori necunoscute x_{ij} care satisfac ecuațiile restrictive dar pentru care forma liniară nu-și atinge neapărat maximul, iar în etapa a doua, soluția inițială este îmbunătățită pînă se ajunge la soluția optimă.

Metoda „diferenței maxime" include în soluția inițială acele rute cu cost minim (sau distanță minimă) care dacă n-ar fi incluse, s-ar cauza pierderi maxime.

Practic, pentru găsirea soluției inițiale se procedează astfel :

Se alcătuiește tabelul — matrice a cheltuielilor unitare $Z = \| Z_{ij} \|$, după care se calculează pentru fiecare linie și fiecare coloană, diferența dintre cel mai mic element al alinierii și cel următor ca mărime. Apoi se caută cea mai mare dintre aceste diferențe. În alinierea în care se realizează acest maxim se caută cel mai mic element ; fie $c_{i_0 j_0}$ acest element și în aceeași poziție din tabloul necunoscutelor se introduce ca valoare a necunoscutelor respective $(x_{i_0 j_0})$ cel mai mic dintre numerele a_{i_0} și b_{i_0}. Celelalte elemente ale alinierii liniei i_0 sau ale coloanei j_0 (din tabelul necunoscutelor) se iau egale cu zero. Prin aceasta se determină elementele unei linii sau ale unei coloane din matricea necunoscutelor și calculele se continuă cu matricea care se obține din C prin eliminarea liniei i_0, respectiv a coloanei j_0. Repetînd acest procedeu se ajunge, pînă la urmă, la determinarea tuturor necunoscutelor, obținîndu-se soluția inițială.

Pentru ilustrarea găsirii soluției inițiale luăm exemplul nostru de mai sus și completăm primul tabel cu diferențele respective dintre elementele Z_{ij} (cheltuielile de transport ale unei tone).

*) Vezi E. Balaș și L. P. Hammer, *Aplicarea metodelor matematice la planificarea transporturilor*. Probleme de automatizare, II, 1960.

Vom avea tabelul 137.

Tabelul 137

Consumator / Producător	C_1	C_2	C_3	C_4	Resurse de material	
P_1	3	2	2	4	70(a_1)	(1)
P_2	1	2	3	4	10(a_2)	(1)
P_3	3	5	2	1	20(a_3)	(1)
Consum	50 (b_1)	25 (b_2)	15 (b_3)	10 (b_4)	100	
	(2)	(3)	(1)	(3)		

În dreptul liniilor și coloanelor s-a specificat în paranteză diferența dintre cel mai mic element din alinierea respectivă și cel următor ca mărime.

Aceste diferențe au, evident, o anumită semnificație economică, ele ne indică pierderea pe care o vom înregistra, dacă pe alinierea respectivă nu vom alege ruta cu costul de transport al unei tone de material cît mai redus.

Întrucît din toate alinierile nu putem alege *numai* rutele cele mai ieftine vor trebui preferate acele rute cu cost minim la care diferența de cost pînă la ruta următoare (în ordinea costurilor) este cea mai mare.

În exemplul nostru diferența cea mai mare se găsește pe coloanele C_2 și C_4; cel mai mic element al coloanei C_2 fiind z_{12} (sau z_{22}) $= 2$, iar al coloanei C_4 fiind $z_{34} = 1$, luăm $x_{34} =$ minim, adică 10 (cantitatea de transport cea mai mică dintre $a_3 = 20$ și $b_4 = 10$); se notează

$$x_{34} = \min\,(a_3,\ b_4) = 10.$$

Prin urmare,

$$x_{14} = x_{24} = 0,$$

altă cantitate care ar trebui transportată nu mai există (centrul de consum C_4 nu mai are nimic de primit). Eliminăm coloana a patra și continuăm calculele pe matricea care se obține.

Tabelul 138

Consumator / Producător	C_1	C_2	C_3	Resurse de material	
P_1	3	2	2	70	(1)
P_2	1	2	3	10	(1)
P_3	3	5	2	10	(1)
Consum	50	25	15	90	
	(2)	(3)	(1)		

Pe această nouă matrice specificăm în dreptul liniilor şi coloanelor producţia respectiv consumul fiecărui centru rămas după prima repartiţie precum şi diferenţele noi dintre cel mai mic element şi elementul următor.

Diferenţa cea mai mare se vede că se găseşte în coloana C_2 (3); cel mai mic element al coloanei fiind z_{22} (sau z_{12}) $= 2$, vom lua $x_{12} = \min(70, 25) = 25$.

Prin urmare, $x_{12} = 25$ şi $x_{22} = x_{32} = 0$.

După eliminarea coloanei C_2 ne rămîne următoarea matrice (tabelul 139):

Tabelul 139

Consumator / Producător	C_1	C_3	Resurse de materiale	
P_1	3	2	45	(1)
P_2	1	3	10	(2)
P_3	3	2	10	(1)
Consum	50	15	65	
	(2)	(1)		

Diferenţa cea mai mare se găseşte pe coloana C_1 (2), cel mai mic element al acestor alinieri fiind z_{21} vom lua $x_{21} = \min(50,10) = 10$. Deci $x_{21} = 10$, $x_{23} = 0$.

Scanned with OKEN Scanner

După eliminarea liniei a doua ne rămîne matricea (tabelul 140).

Tabelul 140

Consumator / Producător	C_1	C_3	Resurse de materiale	
P_1	3	2	45	(1)
P_3	3	2	10	(1)
Consum	40	15	55	
	0	0		

Diferențele cele mai mici fiind pe coloanele C_1 și C_3 vom alege cel mai mic element care este să zicem $z_{33} = 2$; luăm deci $x_{33} = \min(10, 15) = 10$ și eliminăm linia P_3; întrucît $x_{31} = 0$, obținem în sfîrșit ultima matrice (tabelul 141).

Tabelul 141

Consumator / Producător	C_1	C_3	Resurse
P_1	3	2	45
Consum	40	5	45

Se vede că trebuie să luăm $x_{11} = 40$ și $x_{13} = 5$.
Am găsit deci soluția inițială (tabelul 142)

Tabelul 142

$X =$

x_{11}	x_{12}	x_{13}	
x_{21}			
		x_{33}	x_{34}
50	25	15	10

40 3	25 2	5 2		70
10 1				10
		10 2	10 1	20
50	25	15	10	

care satisface ecuațiile restrictive și care ne arată un volum total de cheltuieli de transport egal cu :

$$40.3 + 25.2 + 5.2 + 10.1 + 10.2 + 10.1 =$$
$$= 220 \text{ unități bănești (mii lei).}$$

Această soluție este optimă obținîndu-se din tabelul 142 pentru $x = 5$. În general însă, prin aplicarea procedeului de mai sus se ajunge la început la o soluție apropiată de soluția optimă și apoi prin îmbunătățiri succesive la aceasta din urmă.

Mărirea numărului de date în ipoteza unor probleme mai complicate nu schimbă formularea lor matematică.

Să presupunem astfel că exemplul de mai sus ne interesează reducerea la minimum a cheltuielilor de transport, dar și obținerea unui minim de cheltuieli, ținînd seama în afară de transport și de prețul de cost al producției marfă la fiecare centru A_i.

Dacă notăm cu Z_i, $(i = 1, 2, \ldots, m)$, prețul de cost al produsului considerat la centrul A_i, $(i = 1, 2, \ldots, m)$ și ne propunem să determinăm modul de transport, *astfel ca să obținem cheltuieli totale minime*, atunci vom avea :

— costul total al mărfii produsă în centrul A_i :

$$\sum_{j=1}^{n} Z_i X_{ij} ;$$

— costul transportului produsului de la acest centru la cele n centre de desfacere

$$\sum_{i=1}^{n} C_{ij} X_{ij} .$$

Prin urmare, costul total al mărfii produse și transportate de la centrul A_i este

$$\sum_{j=1}^{n} Z_i X_{ij} + \sum_{j=1}^{n} C_{ij} X_{ij} = \sum_{j=1}^{n} (Z_i + C_{ij}) X_{ij} .$$

După cum se vede, cheltuiala totală se obține adunînd cheltuielile referitoare la fiecare din cele m centre de producție A_i.

Problema de programare se reduce astfel în a face minim volumul total al cheltuielilor.

Trebuie deci să avem

$$\sum_{i=1}^{m} \sum_{j=1}^{n} (Z_i + C_{ij}) X_{ij} = \text{minim.} \qquad \text{(XXIV,4.14)}$$

Este limpede că, în acest caz, numerele X_{ij} trebuie să fie astfel alese încît să realizeze relația de minim (XXIV,4.13) și să îndeplinească în același timp și condițiile (XXIV,4.6), (XXIV,4.7) și (XXIV,4.8).

Scanned with OKEN Scanner

Enunțul — din punctul de vedere matematic — al acestei noi probleme economice, a rămas același ca în problema precedentă.

Introducerea altor condiții ne duce la aceeași concluzie.

Să admitem, de pildă, că la centrul A_i, din cauza unor anumite condiții tehnice și organizatorice, cantitatea de marfă produsă nu poate depăși un anumit maxim d_i.

Va trebui în plus să avem îndeplinită condiția:

$$\sum_{j=1}^{n} X_{ij} \leqslant d_i. \qquad \text{(XXIV, 4.15)}$$

În acest caz nou, problema programării liniare constă în *a face minimă expresia liniară* (*XXIV*,*4.14*) *cu condiția de a avea îndeplinite relațiile* (*XXIV*,*4.6*) (*XXIV*,*4.7*), (*XXIV*,*4.8*) *și* (*XXIV*,*4.15*).

În cazul cînd avem mai multe feluri de produse, ajungem la aceeași formulare matematică.

Trebuie să menționăm, totuși, că în cazul special al problemei transporturilor — ca în exemplul de mai sus, în forma lui inițială — aplicarea metodelor programării liniare poate avea loc cu succes numai în cazul cînd oricare dintre centrele consumatoare poate fi aprovizionat de la oricare dintre centrele producătoare incluse în plan; este vorba, deci, de cazul cînd oricare dintre centrele producătoare incluse în plan și centrele consumatoare, produc și respectiv consumă același produs. În ipoteza unor produse eterogene, pentru care se folosesc unități de măsură diferite, trebuie întocmit cîte un plan optim de transport pentru fiecare produs în parte.

De asemenea, trebuie să mai arătăm că pentru a rezolva o problemă de programare liniară de tipul problemei transporturilor este necesar, evident, ca transportul produsului respectiv să fie planificat de la nivelul unui singur centru coordonator. Această remarcă scoate în evidență faptul că rezolvarea problemei transporturilor — în condițiile formulate mai sus — nu poate fi înfăptuită în țările capitaliste; în aceste țări ea nu poate fi înfăptuită decît în cadrul limitat al relațiilor dintre întreprinderi, izolate, sau în cadrul unui trust. În nici un caz nu este posibilă rezolvarea — prin metodele programării liniare — a problemei transporturilor pe ansamblul economiei naționale.

Scanned with OKEN Scanner

Scanned with OKEN Scanner

Capitolul XXV

Balanţa legăturilor dintre ramuri

1. Consideraţii generale

Pe baza istoricelor hotărîri ale Congresului al III-lea al P.M.R., programul de dezvoltare a economiei naţionale în perspectivă a ţării noastre este orientat în direcţia creşterii intense şi multilaterale a forţelor de producţie în vederea desăvîrşirii construcţiei socialiste şi trecerii treptate la construirea comunismului. Acest program este menit să asigure dezvoltarea armonioasă a tuturor ramurilor economiei naţionale, în conformitate cu cerinţele legii dezvoltării planice proporţionale a economiei naţionale şi a legii economice fundamentale a socialismului, acordîndu-se, în continuare, prioritate industriei grele.

La întocmirea planului de perspectivă şi a planurilor curente, organele de planificare au ţinut şi ţin seamă de cerinţele acestor legi, luînd măsuri din timp pentru a preîntîmpina ivirea oricăror disproporţii în dezvoltarea diferitelor ramuri ale economiei naţionale.

Una dintre cele mai importante metode statistice şi de planificare pentru stabilirea unor proporţii economice juste între producţia mijloacelor de producţie şi producţia obiectelor de consum, între industrie şi agricultură şi diferitele ramuri ale acestora, între veniturile băneşti crescînde ale populaţiei şi masa mărfurilor de consum individual, între acumulare şi consum etc. este — după cum se ştie — *metoda balanţelor*.

Pe baza unui *sistem de balanţe de plan* ale economiei naţionale ca şi a balanţelor de documentare şi de *analiză economico-statistică*, statul nostru stabileşte proporţiile juste în dezvoltarea mereu ascendentă a diferitelor ramuri ale economiei naţionale în condiţiile din etapa actuală a desăvîrşirii construirii socialismului în ţara noastră.

Balanţele constituie, deci, un puternic instrument al planificării economiei naţionale şi al analizei economico-statistice a rezultatelor obţinute.

Proporţiile în dezvoltarea diferitelor ramuri, resursele şi repartiţia lor pe diferite ramuri de producţie şi categorii de produse sînt exprimate

în formă naturală și bănească. Sistemul balanțelor face posibilă compararea resurselor cu nevoile respective și relevarea discordanțelor între diferite ramuri în ceea ce privește nivelul și ritmul lor de dezvoltare; de asemenea, sistemul balanțelor face posibilă descoperirea resurselor suplimentare care se obțin prin economisirea de materii prime și de materiale, ca și prin folosirea mai bună a utilajului.

Analiza prin sistemul balanțelor statistice a resurselor materiale, financiare și de muncă permite înlăturarea rapidă a eventualelor disproporții ce se ivesc în diferitele ramuri ale economiei naționale și asigură totodată posibilitatea manevrării lor elastice.

Din sistemul balanțelor — devenit clasic — elaborat și aplicat cu succes în statistica sovietică și statisticile celorlalte țări ale sistemului economic mondial al socialismului, fac parte în primul rînd, *balanța materială centralizată a producției, consumului și acumulării produsului social* (incluzînd un întreg sistem de balanțe materiale ale diferitelor produse și balanța fondurilor fixe); în al doilea rînd, *balanța resurselor de muncă*, inclusiv *balanța utilizării timpului de muncă*; în al treilea rînd, *balanța financiară* sintetică, oglindind producția, repartiția, repartiția ulterioară și utilizarea finală a venitului național, și în sfîrșit tabel — *balanța sintetică a economiei naționale*, caracterizînd trăsăturile și particularitățile principale ale producției socialiste lărgite în toate momentele acesteia: producția, repartiția, circulația, consumul și acumularea.

În etapa actuală de dezvoltare fără precedent a științei și tehnicii moderne și îndeosebi a tehnicii de calcul electronic, se impune tot mai mult necesitatea ca metodele tehnicii de elaborare a balanțelor economiei naționale să fie ridicate la un nivel calitativ mai înalt folosindu-se în acest scop aparatul modern al matematicii superioare — instrument auxiliar prețios al statisticii. În acest sens, problema centrală care se pune este aceea de a fi elaborată în scurt timp nu o singură variantă a balanței de plan, ci două sau mai multe asemenea variante, alegîndu-se varianta optimă în raport cu condițiile materiale existente.

Acestor exigențe noi răspunde, desigur, aplicarea programării liniare și în domeniul elaborării balanțelor.

După cum vom vedea, aplicarea programării liniare în acest domeniu a dat naștere unei forme noi de balanță denumită *balanța legăturilor dintre ramuri* care este construită după modelul unui tabel-șah și îndeplinește, în esență, aceleași funcții ca și sistemul balanțelor clasice; ea prezintă însă în plus o serie de avantaje incontestabile care o fac să fie un instrument de importanță excepțională în munca de analiză economico-statistică și în munca de planificare.

2. Fundamentarea teoretică a balanței legăturilor dintre ramuri

În concepția economiei politice marxiste, balanța sintetică a economiei naționale, indiferent de forma ei clasică, sau în formă de balanță-șah modernă, împreună cu aparatul ei de calcul matricial și vectorial, ca și întregul sistem de balanțe pe care le întrunește, se fundamentează *pe teoria*

marxistă a reproducției socialiste. Pentru studiul statistic al reproducției socialiste lărgite, un rol călăuzitor, determinant, îl au tezele fundamentale ale teoriei reproducției elaborate de K. Marx.

Schemele lui K. Marx au fost elaborate pentru analiza reproducției capitaliste, totuși, ele conțin o serie de teze fundamentale care-și păstrează întreaga semnificație și pentru economia socialistă. Tezele fundamentale ale teoriei reproducției a lui Marx ca : teza împărțirii producției sociale în producție a mijloacelor de producție și producție a obiectelor de consum, teza creșterii cu precădere a producției mijloacelor de producție în condițiile reproducției lărgite, teza corelației dintre sectoarele I și II, teza formării și destinației fondurilor sociale, teza acumulării ca unica sursă a reproducției lărgite—sînt valabile și în condițiile reproducției socialiste.

În observațiile sale critice asupra cărții lui Buharin „Economia perioadei de tranziție", V.I. Lenin a arătat că formula lui Marx a circulației dintre sectorul I și II rămîne valabilă atît pentru socialism cît și pentru „comunism pur", adică pentru faza a doua a comunismului.

Deosebit de interesantă este remarca pe care o face academicianul S. Strumilin cu privire la tezele lui Marx ca fundament teoretic al sistemului de balanțe în condițiile reproducției socialiste.

Schemele lui Marx — arată Strumilin — „pot fi considerate ca o formă embrionară, o formă inițială sau de pornire, cea mai comprimată variantă a schemelor ce au căpătat o dezvoltare ulterioară în teoria balanțelor socialiste ale economiei naționale, care reprezintă o formă deplin dezvoltată a aceleiași scheme" *). Această reamarcă se referă deopotrivă și la forma nouă a balanței legăturilor dintre ramuri.

În legătură cu fundamentarea teoretică a balanței legăturilor dintre ramuri se mai impun cîteva precizări.

Se știe că în forma și conținutul ei actual, balanța legăturilor dintre ramuri a fost elaborată de V. Leontief **), sub denumirea de „input-output analysis" cu scopul de a fi studiată structura economiei americane pe perioade mai mari — de cîte 10—20 ani.

Fără să intrăm în detalii, se pare că V. Leontief își fundamentează metoda sa de analiză pe teoria echilibrului economic general dezvoltată încă în a doua jumătate a secolului trecut de L. Walras, teorie care domină și astăzi gîndirea economică burgheză. Pe de altă parte, lui V. Leontief nu-i erau străine***) nici schemele lui K. Marx după modelul cărora a fost

---

*) S. Strumilin, *Balans narodnogo hozeaistvakak orudie sotialisticeskogo planirovania*, articol publicat în „Voprosi Economiki" nr. 11, 1954.

**) Ideea de a elabora un studiu privind structura economiei S.U.A. a fost concepută de V. Leontief în 1931. Cercetările sale în această direcție au continuat apoi mai intens în anii celui de-al doilea război mondial cînd apar cunoscutele sale lucrări asupra structurii economiei americane, una din ele cuprinzînd perioada 1919/929 iar a doua perioadă 1929/939. După război V. Leontief a adîncit cercetările sale, ocupîndu-se și de latura dinamică a modelului său input-output.

***) În prefața traducerii în limba rusă a lucrării lui Oskar Lange, *Nek torlie soobrajenia po analizu balansa zatrat i vipuska producții*, academicianul V. Nemcinov atrage atenția că V. Leontief a lucrat pînă în anul 1926 la gosplanul U.R.S.S. și că deci cunoștea schema reproducției

elaborată prima balanță-șah pentru studiul economiei naționale a U.R.S.S. pe anul 1923/1924. Se poate ușor demonstra, examinînd această primă lucrare cunoscută în literatura economică cu privire la studiul unui sistem economic în ansamblu prin metoda balanței-șah, că prin concepția ei, schema balanței U.R.S.S. pe anul 1923/1924, are o serie întreagă de elemente care au putut exercita influență directă asupra modelului input-output conceput mai tîrziu de V. Leontief.

Faptul că Leontief a cunoscut schema lui Marx rezultă evident din însuși modelul său input-output închis care constituie o schemă tipică de reproducție simplă a lui Marx.

Trebuie să arătăm însă — indiferent de concepția teoretică a lui V. Leontief — că el are meritul incontestabil de a fi elaborat un model practic pentru cunoașterea concretă a structurii economiei americane. Desigur că perspectivele aplicării acestui model în economia capitalistă sînt limitate știut fiind că el nu este decît un caz particular de aplicare a programării liniare. Or, în condițiile economiei capitaliste, am văzut cadrul îngust în care aceasta își poate găsi aplicații utile.

Este interesant de observat că în bogata literatură apărută în țările capitaliste pe tema metodei de analiză input-output se atrage adeseori atenția că ea constituie un caz particular al programării liniare în care nu se pune problema alegerii unei variante alternative sau problema folosirii optime a resurselor *). Aceasta constituie o recunoaștere a faptului că problema soluției optime în folosirea resurselor, pusă pe ansamblul economiei unui stat capitalist nu se poate rezolva.

Rezultă, deci, că numai în condițiile economiei socialiste în care acționează legea dezvoltării planice proporționale a economiei naționale și legea economică fundamentală a socialismului, metoda balanței legăturilor dintre ramuri își poate găsi o „valorificare" maximă.

3. Schema balanței legăturilor dintre ramuri **)

Vom expune pe scurt conținutul și esența acestei scheme.

Dacă notăm cu:

X_i, producția globală a ramurii i, $(i = 1, 2, \ldots, n)$;

x_{ij}, cantitatea produsului creat în ramura i însă transferată pentru consum productiv în ramura j, $(j = 1, 2, \ldots, n)$, materie primă, combustibil etc.;

---

lărgite a lui Marx și lucrările care se executau atunci pentru întocmirea balanței economiei naționale a U.R.S.S. pe anii 1923/924. De altfel în anul 1925 în revista „Planovoie hoziaistvo" nr. 12, a apărut recenzia lui V. Leontiev despre prima balanță a economiei naționale a U.R.S.S.

*) Vezi în acest sens remarca pe care o face I. P. Gherciuc, pe marginea articolului lui G. Marton apărut parțial, tradus în limba rusă, într-o broșură de material documentar editată de Academia de Științe a U.R.S.S., filiala Siberia, 1959.

**) Folosirea acestei terminologii — deși nu tocmai exactă — o socotim cea mai potrivită. Prin termenul de „legături dintre ramuri" trebuie să se înțeleagă nu numai legăturile dintre ramurile economiei naționale ci și cele legate de activitatea în domeniul investițiilor, comerțului exterior, sau legăturile interregionale etc.

x_i, produsul „final” *) creat în ramura i, adică acea parte a produsului global X_i care nu intră în consumul altor ramuri pentru producția curentă. Produsul final poate fi folosit pentru consumul individual, consumul colectiv, pentru acumulare sau pentru export.

Pe baza acestor indicații vom putea scrie sintetic următoarea relație :

$$X_i = \sum_{j=1}^{n} x_{ij} + Y_i , \qquad (i = 1, 2, \ldots, n) \quad \text{(XXV, 3.1)}$$

Ea exprimă de fapt un sistem întreg de relații de balanță între producția și repartiția produsului social pe ramuri ale economiei naționale care poate fi ilustrat prin următorul tabel în forma tablei de șah :

$$\begin{array}{c|c|c} X_1 & x_{11}\ x_{12} \ldots x_{1j} \ldots x_{1n} & Y_1 \\ X_2 & x_{21}\ x_{22} \ldots x_{2j} \ldots x_{2n} & Y_2 \\ \vdots & \ldots\ldots\ldots\ldots & \vdots \\ X_i & x_{i1}\ x_{i2} \ldots x_{ij} \ldots x_{in} & Y_i \\ \vdots & \ldots\ldots\ldots\ldots & \vdots \\ X_n & x_{n1}\ x_{n2} \ldots x_{nj} \ldots x_{nn} & Y_n \end{array} \qquad \text{(XXV, 3.2)}$$

Mărimile x_{ij} din matricea pătrată aflată în centrul acestui tabel reprezintă tocmai legăturile cantitative reciproce între diferitele ramuri — în prima coloană din strînga matricei avem producția globală, iar în coloana din dreapta — producția „finală”.

În sistemul balanței-șah de mai sus poate fi introdusă și balanța forței de muncă folosită în procesul de producție. Notînd cu S_0 volumul total al forței de muncă ce poate fi folosită, cu s_j forța de muncă folosită în ramura j și cu s_0 forța de muncă nefolosită în producția materială cerută (de pildă, cea angajată în sfera neproductivă sau care reprezintă rezervă nefolosită).

Cu aceste notații balanța forței de muncă poate fi exprimată astfel :

$$S_0 = s_0 + \sum_{j=1}^{n} s_j . \qquad \text{(XXV, 3.3)}$$

*) Această parte a produsului global al ramurii i constituie un element important al producției nete (venitului național) al întregii economii naționale, valoarea ei nu este însă egală cu valoarea nou creată în ramura i dată.

Introducînd și această balanță în balanța de mai sus, vom avea următorul tabel-șah :

$$\begin{array}{c|c|c}
X_1 & x_{11}\ x_{12}\ \dots\ x_{1j}\ \dots\ x_{1n} & Y_1 \\
X_2 & x_{21}\ x_{22}\ \dots\ x_{2j}\ \dots\ x_{2n} & Y_2 \\
\vdots & \dots\dots\dots\dots & \vdots \\
X_i & x_{i1}\ x_{i2}\ \dots\ x_{ij}\ \dots\ x_{in} & Y_i \\
\vdots & \dots\dots\dots\dots & \vdots \\
X_n & x_{n1}\ x_{n2}\ \dots\ x_{nj}\ \dots\ x_{nn} & Y_n \\
\hline
S_0 & s_{01}\ s_{02}\ \dots\ s_{0j}\ \dots\ s_{0n} & s_0 \\
 & C_1\ C_2\ \dots\ C_j\ \dots\ C_n &
\end{array} \qquad \text{(XXV, 3.4)}$$

Pe ultima linie apar valorile C_j, $(j = 1, 2, \dots, n)$, care reprezintă suma cheltuielilor de producție în ramura j. Este evident că însumarea pe coloane este posibilă numai în ipoteza în care elementele tabelului (XXV, 3.4) sînt date în expresie bănească; în caz contrar, însumarea elementelor matricei se poate face numai pe linii orizontale, X_i reprezentînd în acest caz o producție omogenă oarecare, exprimată în anumite unități de măsură naturale (tone, metri etc.).

În ipoteza unui tabel-șah valoric avem relația :

$$C_j = s_{0j} + \sum_{i=1}^{n} x_{ij}, \qquad (j = 1, 2, \dots, n) \qquad \text{(XXV, 3.5)}$$

care exprimă cheltuielile de producție (mijloacele de producție plus salariile) în ramura j.

Tabelul-șah (XXV, 3.4) — în expresie valorică — poate fi completat și cu acumulările din ramura j, în care caz valorile însumate pe coloane sînt egale cu cele însumate pe linii orizontale.

Notînd cu m_j acumulările în ramura j, vom avea următorul tabel-balanță complet al legăturilor dintre ramuri, de tipul unui tabel deschis*) :

$$\begin{array}{c|c|c}
X_1 & x_{11}\ x_{12}\ \dots\ x_{1j}\ \dots\ x_{1n} & Y_1 \\
X_2 & x_{21}\ x_{22}\ \dots\ x_{2j}\ \dots\ x_{2n} & Y_2 \\
\vdots & \dots\dots\dots\dots & \vdots \\
X_i & x_{i1}\ x_{i2}\ \dots\ x_{ij}\ \dots\ x_{in} & Y_i \\
\vdots & \dots\dots\dots\dots & \vdots \\
X_n & x_{n1}\ x_{n2}\ \dots\ x_{nj}\ \dots\ x_{nn} & Y_n \\
\hline
S_0 & s_{01}\ s_{02}\ \dots\ s_{0j}\ \dots\ s_{0n} & s_0 \\
 & m_1\ m_2\ \dots\ m_j\ \dots\ m_n & \\
 & X_1\ X_2\ \dots\ X_j\ \dots\ X_n &
\end{array} \qquad \text{(XXV,3.6)}$$

*) Un tabel închis se referă numai la matricea $|\ x_{ij}\ |$ fără celelalte elemente.

Se vede din tabel că producția globală X_j poate fi obținută însumînd fie valorile de pe liniile orizontale fie cele de pe coloane.

Avem, egalitatea:

$$X_j = \sum_{j=1}^{n} x_{ij} + Y_i = s_{0i} + \sum_{j=1}^{n} x_{ji} + m_i, \qquad (i = 1, 2, \ldots, n) \qquad \text{(XXV, 3.7)}$$

Pentru balanța legăturilor dintre ramuri în expresie naturală se obișnuiește notația: Q_i pentru produsul ramurii i; Q_{ij} pentru partea produsului Q_i ce se consumă în ramura j (care exprimă tocmai relația cantitativă dintre ramuri); Q_0 pentru totalul forței de muncă folosite (om-ore, om-zile etc.), iar Q_{0i} și Q_0 pentru forța de muncă folosită în ramura i și respectiv — pentru partea de muncă nefolosită în sfera producției materiale.

Pentru această balanță avem evident, legătura de balanță pe rînduri orizontale, exprimate prin relația:

$$Q_i = \sum_j q_{ij} + q_i, \qquad (i, j = 1, 2, \ldots, n), \qquad \text{(XXV, 3.8)}$$

q_i reprezentînd produsul final (neconsumat în procesul de producție).

4. Coeficienții tehnici și unele aplicații ale balanței legăturilor dintre ramuri

De o importanță deosebită pentru studiul condițiilor tehnologice ale producției este stabilirea așa-numiților *coeficienți tehnici* ai consumurilor de materiale și de forță de muncă, adică acele consumuri care reprezintă consumurile productive ale ramurii considerate.

În țările lagărului socialist acești coeficienți sînt planificați sub forma *normelor tehnice*, adică a *consumurilor specifice* de materiale, de forță de muncă și de alte elemente. Dacă asemenea norme tehnice n-au fost planificate pentru unele ramuri de producție, ele se pot calcula cu o anumită aproximație pe baza balanței statistice a legăturilor dintre ramuri, se fac în acest scop rapoartele

$$\frac{q_{ij}}{Q_j} = a_{ij}. \qquad \text{(XXV, 4.1)}$$

Mărimile a_{ij} reprezintă tocmai coeficienții tehnici în chestiune, arătînd cît anume se consumă din produsul ramurii i pentru producția unei unități de produs al ramurii j.

Dacă producția este exprimată în unități valorice, mărimile corespunzătoare:

$$\frac{x_{ij}}{X_j} = a_{ij} \qquad \text{(XXV, 4.2)}$$

vor reprezenta coeficienții cheltuielilor dorite de producție.

Din relațiile (XXV, 4.1) și (XXV, 4.2) deducem egalitățile:

$$q_{ij} = a_{ij} Q_j,$$
$$x_{ij} = a_{ij} X_j.$$

Rezultă deci că putem scrie următoarele legături de balanță:

$$Q_i = \sum_j a_{ij} Q_j + q_i, \qquad (i = 1, \ldots, n)$$

sau respectiv (XXV, 4.3)

$$X_i = \sum_j a_{ij} X_j + Y_i, \qquad (i = 1, \ldots, n)$$

Referindu-ne la balanța-șah în expresie valorică și ținînd seama că avem egalitatea $X_i = a_{ij} X_j$, putem scrie următorul sistem de ecuații cu n necunoscute:

$$\begin{aligned} X_1 &= a_{11} X_1 + a_{12} X_2 + \ldots + a_{1n} X_n + Y_1 \\ X_2 &= a_{21} X_1 + a_{22} X_2 + \ldots + a_{2n} X_n + Y_2 \\ &\ldots\ldots \quad \ldots\ldots\ldots\ldots\ldots \\ X_n &= a_{n1} X_1 + a_{n2} X_2 + \ldots + a_{nn} X_n + Y_n \end{aligned} \qquad \text{(XXV, 4.4)}$$

care sub formă matricială se scrie:

$$X = AX + Y \qquad \text{(XXV, 4.5)}$$

în care

$$X = \begin{Vmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{Vmatrix}$$

este vectorul produsului total al fiecărei ramuri,

$$Y = \begin{Vmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{Vmatrix}$$

este vectorul produsului „final" al fiecărei ramuri, iar A este matricea coeficienților tehnici:

$$A = \begin{Vmatrix} a_{11} & a_{12} & \ldots & a_{1n} \\ a_{21} & a_{22} & \ldots & a_{2n} \\ \ldots & \ldots & \ldots & \ldots \\ a_{n1} & a_{n2} & \ldots & a_{nn} \end{Vmatrix}.$$

Soluția sistemului matricial (XXV, 4.5) se obține considerînd matricea unitate

$$E = \begin{Vmatrix} 1 & 0 & 0 & \dots & 0 \\ 0 & 1 & 0 & \dots & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & \cdot & \cdot & \cdot & 1 \end{Vmatrix}$$

apoi matricea $[E - A]$ și inversa ei $[E - A]^{-1}$

Obținem astfel:

$$\begin{aligned} Y &= (E - A)\, X \\ X &= (E - A)^{-1}\, Y \end{aligned} \qquad \text{(XXV, 4.6)}$$

Sistemul (XXV, 4.5) are n ecuații cu n necunoscute X_i, $(i = 1, 2, \dots, n)$. În general acest sistem este compatibil. Condiția este ca determinantul coeficienților să fie diferit de zero.

Se știe că avem:

$$EA = AE = A, \qquad EX = XE = X.$$

Relația (XXV, 4.5), poate deci fi scrisă:

$$\begin{aligned} EX &= AX + Y, \\ EX - AX &= Y, \\ (E - A)X &= Y. \end{aligned}$$

Înmulțim la stînga ambii membri ai acestei ecuații cu $(E - A)^{-1}$:

$$(E - A)^{-1}(E - A)\, X = (E - A)^{-1}\, Y.$$

Prin definiție:

$$(E - A)^{-1} \cdot (E - A) = E.$$

Prin urmare, rezultă că:

$$EX = (E - A)^{-1}\, Y$$

sau

$$X = (E - A)^{-1}\, Y$$

adică tocmai ecuația a doua din relația de mai sus (XXV, 4.6).

A doua ecuație matricială din (XXV, 4.6) scoate în evidență faptul că matricea inversă $(E - A)^{-1}$ ne ajută să rezolvăm problema consumului în ramura j sub un alt aspect și anume acela al consumului *total* al produsului X_i pentru producerea unei unități de produs al ramurii j. Consumurile totale sau respectiv, cheltuielile totale (cînd balanța este întocmită în expresie bănească) sînt date de elementele matricei inverse $(E - A)^{-1}$ prin categoriile de consumuri (cheltuieli) totale, înțelegîndu-se în afară de cele ce au loc *direct* în procesul de producție al ramurii j și cele care

au loc în întreaga economie națională pentru producerea unei unități de produs în ramura j (consumuri — cheltuieli, *indirecte*).

În același mod putem stabili un sistem de relații între prețurile medii și componentele lor.

Să considerăm din nou relația (XXV, 3.7):

$$X_i = \sum_{j=1}^{n} x_{ij} + s_{0i} + m_i .$$

Notînd cu P_i prețul mediu unitar în ramura i, avem relațiile:

$$\begin{aligned} X_i &= Q_i P_i, && (i = 1, 2, \ldots, n), \\ x_{ij} &= q_{ij} \cdot P_i, && (i,j = 1, 2, \ldots, n). \end{aligned} \qquad \text{(XXV, 4.7)}$$

Pentru uniformitate introducem și relațiile:

$$\begin{aligned} s_{0i} &= Q_i \bar{s}_{0i}, && (i = 1, 2, \ldots, n) \\ m_i &= Q_i \bar{m}_i, && (i = 1, 2, \ldots, n) \end{aligned} \qquad \text{(XXV, 4.8)}$$

în care $\bar{s}_{0i}$ reprezintă salariul mediu pe unitatea de produs în ramura i și m_i, acumularea medie în ramura i, de asemenea pe unitate de produs.

Introducînd în relația (XXV, 3.7) valorile din relațiile (XXV, 4.7) și (XXV, 4.8), obținem:

$$Q_i P_i = \sum_{j=1}^{n} q_{ji} P_j + Q_i \bar{s}_{0i} + Q_i \bar{m}_i, \qquad (i = 1, 2, \ldots, n)$$

și apoi împărțind ambele părți ale egalității la Q_i, avem:

$$P_i = \sum_{j=1}^{n} \frac{q_{ji}}{Q_i} P_j + \bar{s}_{0i} + \bar{m}_i, \qquad (i = 1, 2, \ldots, n).$$

Ținînd seama de relațiile (XXV, 4.1), obținem mai departe:

$$P_i = \sum_{j=1}^{n} a_{ji} P_j + \bar{s}_{0i} + \bar{m}_i, \qquad (i = 1, 2, \ldots, n) \qquad \text{(XXV, 4.9)}$$

Egalitățile (XXV, 4.9) reprezintă sistemul fundamental de relații dintre prețurile medii p_i și componentele lor.

Vom nota acum prin:

$$A' = \begin{Vmatrix} a_{11} & a_{21} & \ldots & a_{n1} \\ a_{12} & a_{22} & \ldots & a_{n2} \\ \ldots & \ldots & \ldots & \ldots \\ a_{1n} & a_{2n} & \ldots & a_{nn} \end{Vmatrix}$$

transpusa matricii A, și prin $\bar{s}_0$, $\bar{m}$ și P respectiv matricile:

$$\bar{s}_0 = \begin{Vmatrix} \bar{s}_{01} \\ \bar{s}_{02} \\ \vdots \\ \bar{s}_{0n} \end{Vmatrix}, \quad \bar{m} = \begin{Vmatrix} \bar{m}_1 \\ \bar{m}_2 \\ \vdots \\ \bar{m}_n \end{Vmatrix}, \quad P = \begin{Vmatrix} P_1 \\ P_2 \\ \vdots \\ P_n \end{Vmatrix};$$

sistemul (XXV, 4.9) se poate scrie matricial:

$$P = A'P + \bar{s}_0 + \bar{m}\,. \qquad \text{(XXV, 4.9')}$$

Soluția acestui sistem este:

$$P = [(E-A)']^{-1} \cdot [\bar{s}_0 + \bar{m}] \qquad \text{(XXV, 4.10)}$$

Sistemul (XXV, 4.10) reprezintă n ecuații care leagă între ele valorile P_i, $\bar{s}_{0i}$, și $\bar{m}_i$, al căror număr este egal cu 3 n. El ne permite ca în baza matricei fundamentale A să rezolvăm numeroase probleme de planificare în domeniul prețurilor, salariilor și acumulării.

În cazul nostru am presupus cunoscute $\bar{s}_{0i}$ și $\bar{m}_i$ și în consecință am stabilit prețurile unitare în toate ramurile economiei naționale.

Dacă presupunem că salariile se schimbă la unele ramuri, atunci tot din sistemul (XXV, 4.10), găsim modificările ce se impun a fi făcute prețurilor în toate ramurile (în ipoteza că acumulările rămîn la același nivel). Sau putem să ne punem problema de a mări sau micșora acumularea în unele ramuri, paralel cu creșterea salariilor în alte ramuri, în care caz, tot sistemul (XXV, 4.10) ne dă soluția pentru prețurile medii în toate ramurile economiei naționale.

Este de remarcat că rezolvarea acestor probleme se bazează pe calculul matricei:

$$[(E - A)']^{-1},$$

după cum toate problemele legate de calculul balanței producției și consumului se rezolvă cu ajutorul matricei:

$$(E - A)^{-1}.$$

Dificultatea principală constă, însă, în calculul acestor matrici inverse; ele necesită calcule numeroase din cauza numărului mare de elemente pe care le conțin. Pentru a simplifica calculele se poate folosi o singură operație de inversare a matricei.

Într-adevăr, să aplicăm operația de transpunere în ambii membri ai relației (XXV, 4.10):

$$P' = \{[(E - A)']^{-1} (\bar{s}_0 + \bar{m})\}'. \qquad \text{(XXV, 4.11)}$$

Or, dacă avem două matrici C și D, știm că

$$(C\,D)' = D'C',$$

$$(C')^{-1} = (C^{-1})'\,;\ (C')' = C.$$

Ținînd seama de aceste reguli, produsul de matrici din membrul al doilea al relației (XXV, 4.11) devine succesiv:

$$\{[(E - A)']^{-1} (\bar{s}_0 + \bar{m})\}' = [\bar{s}_0 + \bar{m}\,]'\{[(E - A)']^{-1}\}' =$$

$$= [(\bar{s}_0' + \bar{m}')\,(E - A)'']^{-1} = [(\bar{s}_0' + \bar{m}')\,(E - A)]^{-1}$$

și atunci sistemul (XXV, 4.10) se poate scrie sub forma echivalentă:

$$P' = (\varepsilon_0' + m')(E - A)^{-1}. \qquad \text{(XXV, 4.12)}$$

Folosim, deci, aceeași matrice inversă ca cea de la (XXV, 4.6) cu deosebire că în relația (XXV, 3.12) matricea inversă $(E - A)^{-1}$ înmulțește la dreapta vectorul produsului „final".

*

Balanța legăturilor dintre ramuri constituie un instrument deosebit de prețios pentru munca de analiză și planificare, îndeosebi, — la nivelul economiei naționale. În sistemul economiei socialiste ea își poate găsi numeroase domenii de aplicare. În afară de domeniul balanțelor de producție și repartiție a produsului social total sau al prețurilor, ea poate fi aplicată cu succes, de pildă, la rezolvarea problemelor de eficiență economică a investițiilor, la analiza eficienței economice a comerțului exterior, la analiza cheltuielilor de producție, sau a consumului unor produse izolate și în multe alte probleme. Cadrul lucrării nu ne permite, desigur, să abordăm toate aceste probleme, fiecare cu specificul ei, însă atragem atenția cititorilor că din punct de vedere matematic toate aceste probleme se rezolvă după același principiu, expus mai sus.

Scanned with OKEN Scanner

Capitolul XXVI

Diverse generalizări ale metodei programării liniare

1. Caracterul statistic al programării liniare

Să considerăm următorul exemplu:

O gospodărie agricolă de stat dispune de 1000 ha teren arabil și de un fond de 2 400 000 lei. Ea poate însămînța trei culturi diferite (A, B și C) în următoarele condiții:

1. — cultura A produce în medie 1 800 kg la ha;
   — „ B „ „ „ 2 400 „ „ „ ;
   — „ C „ „ „ 2 000 „ „ „ ;
2. — cheltuielile pentru obținerea culturii A sînt de 2 000 lei/ha;
   — „ „ „ „ „ B „ „ 3 000 „ ;
   — „ „ „ „ „ C „ „ 2 500 „ ;
3. — un kg de produs din cultura A se vinde cu 2 lei ;
   — „ „ „ „ „ „ „ B „ „ „ 2,58 „ ;
   — „ „ „ „ „ „ „ C „ „ „ 2,50 „ ;

Pentru a-și ridica la nivel mai înalt rentabilitatea, această gospodărie și-a propus să stabilească în așa mod numărul de ha la care să se aplice fiecare din culturile A, B, C, încît să se realizeze un venit maxim posibil, dat fiind că nu există nici un alt considerent de ordin economic în determinarea suprafețelor care vor fi destinate uneia din aceste culturi.

Însemnînd prin x numărul de ha cultivate cu cultura A, prin y, numărul de ha cultivate cu cultura B și prin z, numărul de ha cultivate cu cultura C; trebuie să avem în condițiile date:

$$\left.\begin{aligned} x + y + z &= 1\,000 \\ x \geqslant 0,\ y &\geqslant 0,\ z \geqslant 0 \\ 2\,000\,x + 3\,000\,y + 2\,500\,z &\leqslant 2\,400\,000 \end{aligned}\right\} \quad \text{(XXVI, 1.1)}$$

Beneficiul total va fi dat de expresia liniară:

$$2.\ 1\,800\,x+2{,}58.2\,400\,y+2{,}5.2\,000\,z-2\,000\,x-3\,000\,y-2\,500\,z = \\ = 1\,600\,x+3\,200\,y+2\,500\,z. \qquad \text{(XXVI, 1.2)}$$

Problema de programare liniară care se pune constă în a găsi maximul funcției (XXVI, 1.2) de mai sus; trebuie să avem:

$$1\,600\,x+3\,200^{*})\,y+2\,500\,z = \text{maxim} \qquad \text{(XXVI, 1.3)}$$

fiind îndeplinite și condițiile (XXVI, 1.1).

Vom reduce cele trei necunoscute la 2 și anume x și y, folosind relația:

$$z = 1\,000 - x - y$$

care rezultă din prima din relațiile (XXVI, 1.1).

Efectuînd această substituție în fiecare din relațiile (XXVI, 1.1), datele problemei se simplifică. Știm că avem:

$$\left.\begin{aligned} x - y &\geqslant 200,\\ x + y &\leqslant 1000,\\ x &\geqslant 0,\\ y &\geqslant 0. \end{aligned}\right\} \qquad \text{(XXVI, 1.4)}$$

Problema se pune de a determina astfel valorile lui x și y ca să avem

$$-9x+7y = \text{maxim}. \qquad \text{(XXVI, 1.5)}$$

Grafic, soluția problemei rezultă imediat (vezi fig. 53).

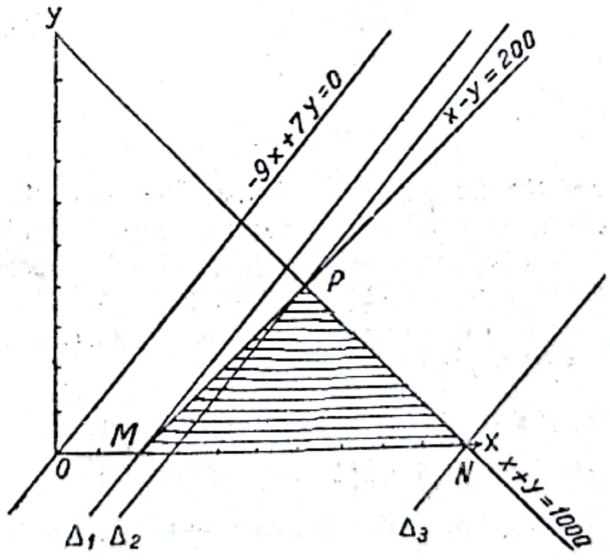


Fig. 53

Figurăm în planul $x0y$, dreptele $x = 0$, $y = 0$, $x-y = 200$, $x+y = 1\,000$. Punctele din plan care satisfac inegalitățile (XXVI, 1.4)

*) Numărul s-a rotunjit în plus cu 8 sutimi pentru ușurarea calculelor.

se găsesc în regiunea haşurată, care reprezintă interiorul triunghiului MNP ale cărui vîrfuri au respectiv coordonatele: $M(200,0)$, $N(1000,0)$, $P(600, 400)$.

Se ştie că maximul sau minimul funcţiei liniare cade într-unul din vîrfurile regiunii haşurate, adică într-unul din punctele M, N, P. Înlocuind coordonatele acestor puncte în funcţia (XXVI, 1.5) se găseşte că maximul corespunde punctului $M(200,0)$, pentru care avem:

$$x = 200,\ y = 0,\quad z = 800.$$

O altă metodă pentru determinarea punctului de maxim constă în ducerea prin origine a dreptei $-9x + 7y = 0$ şi prin trasarea unor drepte paralele cu această dreaptă, Δ_1, Δ_2, Δ_3, care trec respectiv prin vîrfurile triunghiului MNP. Cea mai apropiată de origine dintre ele trece prin punctul de maxim. Se vede pe figură că aceasta este dreapta Δ_1 care trece prin M şi că deci M este punctul căutat. Prin urmare, venitul maxim corespunde folosirii a 200 ha pentru cultura A şi 800 ha pentru cultura C. El este egal cu 2 320 000 lei. Orice altă soluţie conduce la un venit mai mic.

Soluţia problemei nu poate fi considerată satisfăcătoare din punct de vedere economic. La întocmirea planului de cultură cunoaştem producţia medie la ha, însă nu sîntem siguri că în realitate, după strîngerea recoltei, vom obţine tocmai această medie. Singura apreciere pe care o putem face este că producţia recoltată, cu o mare probabilitate, va fi cuprinsă într-un anumit interval.

Referitor la cele trei culturi, vom presupune deci: producţia la ha se distribuie după legea normală cu următoarele valori tipice:

Cultura	*Producţia medie*	*Dispersia*
A	1 600	180
B	3 200	360
C	2 500	250

În acest caz, producţia culturii A este $1\,600 + 180\,k$, unde k este în imensa majoritate a cazurilor un număr cuprins între -3 şi 3 (în $99{,}7\%$ din cazuri). La fel pentru culturile B şi C, pentru care avem respectiv producţiile $3200 + 360\,k'$ şi $2500 + 250\,k''$.

Condiţiile (1), (2) rămîn aceleaşi, iar funcţia $F(x, y, z)$, care trebuie să fie maximă, devine în acest caz:

$$F(x,y,z) = (1600 + 180\,k)\,x + (3200 + 360\,k')\,y + (2500 + 250\,k'')\,z \qquad \text{(XXVI, 1.6)}$$

Maximul lui $F(x, y, z)$ cade tot într-unul din vîrfurile triunghiului MNP.

Calculînd valorile funcţiei $F(x, y, z)$ în fiecare din punctele M, P, N, obţinem următoarele valori:

$F = 2\,320\,000 + 36\,000\,k + 200\,000\,k''$, în punctul M.
$F = 2\,240\,000 + 108\,000\,k + 144\,000\,k'$, „ „ P.
$F = 1\,600\,000 + 180\,000\,k$ „ „ N.

Scanned with OKEN Scanner

de unde rezultă că, pentru ca maximul să cadă în punctul M, trebuie să fie îndeplinite condițiile:

$$\left.\begin{aligned} 10 - 9k - 18k' + 25k'' > 0, \\ 90 - 18k + 25k'' > 0. \end{aligned}\right\} \qquad \text{(XXVI, 1.7)}$$

Maximul cade în P, dacă sînt îndeplinite condițille:

$$\left.\begin{aligned} 10 - 9k - 18k' + 25k'' < 0, \\ 80 - 9k + 18k' > 0. \end{aligned}\right\} \qquad \text{(XXVI, 1.8)}$$

În sfîrșit, poziția de maxim corespunde punctului N, dacă:

$$\begin{aligned} 90 - 18k + 25k'' < 0, \\ 80 - 9k + 18k' < 0. \end{aligned} \qquad \text{(XXVI, 1.9)}$$

Numerele k, k', k'' urmează fiecare o lege normală cu media zero și dispersia 1. Dacă presupunem că aceste legi sînt independente, un calcul simplu de probabilități ne arată că:

$$\text{Prob. } \{10 - 9k - 18k' + 25k'' > 0\} = 0{,}5120, \qquad \text{(XXVI, 1.10)}$$

$$\text{Prob. } \{90 - 18k + 25k'' > 0\} = 0{,}9987, \qquad \text{(XXVI, 1.10')}$$

$$\text{Prob. } \{80 - 9k + 18k' > 0\} = 0{,}9999. \qquad \text{(XXVI, 1.10'')}$$

Formulele (XXVI, 1.10) ne arată că inegalitățile (XXVI, 1.9) sînt îndeplinite cu o probabilitate mai mică decît 10^{-6}. În mai puțin decît un caz la un milion se va putea ca venitul maxim să corespundă punctului N (1000, 0, 0). Din punct de vedere practic, înlăturăm această soluție.

Rămîne deci să examinăm numai inegalitățile (XXVI, 1.7) și (XXVI, 1.8) care corespund punctelor M și P.

Ținînd seama de faptul că ultimile două probabilități din (XXVI, 1.10) sînt mari, putem enunța următorul rezultat: cu o probabilitate egală cu 0,51 soluția optimă este $x = 200$, $y = 0$, $z = 800$, și cu o probabilitate egală cu 0,49, soluția optimă este $x = 600$, $y = 400$, $z = 0$.

Rezultatul pe care l-am obținut pune în evidență caracterul statistic al metodei programării liniare. Am arătat mai sus (metoda grafică) că venitul maxim se obține dacă se aplică cultura A la 200 ha și cultura C la 800 ha. Completarea de natură statistică a acestui rezultat ne arată că numai în 51% din cazuri venitul maxim corespunde acestei soluții și că în 49% din cazuri venitul maxim se va obține dacă se vor rezerva 600 ha culturii A și 400 ha culturii B. Rezultatul se explică prin faptul că producția la hectar nu este dată de un număr precizat de mai înainte, ci de un număr care variază, în imensa majoritate a cazurilor, într-un interval cunoscut.

Alegerea cu o anumită probabilitate a soluției optime se datorește prin urmare faptului că nu putem prevedea de la început valoarea exactă a producției pe hectar.

Este de observat că problema a fost rezolvată în ipoteza că producțiile A, B, C, sînt independente între ele. În cazul cînd se dau legile

exacte de probabilitate a acestor producții, soluția se obține în același mod, astfel încît ipoteza făcută nu conduce decît la eventuale simplificări de calcul.

Notăm prin α producția pentru cultura A, prin β producția pentru cultura B și prin γ producția pentru cultura C. Fie $f(\alpha, \beta, \gamma)$ densitatea de probabilitate pentru variabilele α, β, γ.

Este evident că avem:

$$f(\alpha, \beta, \gamma) \geqslant 0,$$

$$\iiint_T f(\alpha, \beta, \gamma)\, d\alpha\, d\beta\, d\gamma = 1$$

integrala fiind întinsă la spațiul T: $\alpha \geqslant 0$, $\beta \geqslant 0$, $\gamma \geqslant 0$.

Funcția $F(x, y, z)$ al cărei maxim îl căutăm este în acest caz:

$$F(x, y, z) = (2\alpha - 2\,000)x + (2{,}58\beta - 3\,000)y + (2{,}5\,\gamma - 2\,500)z$$

$F(x, y, z)$ devine maximă tot într-unul din punctele

$M(200, 0, 800)$, $P\ (600, 400, 0)$, $N\ (1000, 0, 0)$.

În spațiul valorilor α, β, γ, $F(x, y, z)$ atinge valoarea maximă în punctul M cînd (α, β, γ) este situat în regiunea T_1, determinată prin relațiile:

$$2\alpha + 2{,}58\beta \leqslant 5\gamma,$$
$$4\alpha + 1000 \leqslant 5\gamma.$$

Maximul corespunde punctului P, în regiunea T_2, determinată prin egalitățile:

$$2\alpha + 2{,}58\beta \geqslant 5\gamma,$$
$$2\alpha + 1000 \leqslant 2{,}58\beta.$$

În sfîrșit, maximul este dat de punctul N în regiunea T_3, definită prin:

$$2{,}58\beta \leqslant 2\alpha + 1000,$$
$$5\gamma \leqslant 4\alpha + 1000.$$

Probabilitatea ca soluția optimă să corespundă punctului M este ată de integrala:

$$\iiint_{T_1} f(\alpha, \beta, \gamma)\, d\alpha\, d\beta\, d\gamma.$$

În mod analog sînt determinate probabilitățile referitoare la soluțiile date de punctele P și N.

Soluționarea problemei se face ținînd seama de valoarea medie a funcției $F(x, y, z)$ al cărei maxim îl căutăm.

În regiunea T_1, valoarea funcției $F(x, y, z)$ este:

$$400\ \alpha + 2000\ \gamma - 2\,400\,000.$$

În regiunile T_2 și T_3 valorile lui $F(x, y, z)$ sînt respectiv

$$1200\ \alpha + 1032\ \beta - 2\ 400\ 000,$$
$$2000\ \alpha - 2\ 000\ 000.$$

Avem prin urmare de calculat integralele:

$$\iiint_{T_1} (400\ \alpha + 2000\ \gamma - 2\ 400\ 000)\quad f(\alpha, \beta, \gamma)\, d\alpha\, d\beta\, d\gamma,$$

$$\iiint_{T_2} (1200\ \alpha + 1032\ \beta - 2\ 400\ 000)\quad f(\alpha, \beta, \gamma)\, d\alpha\, d\beta\, d\gamma,$$

$$\iiint_{T_3} (2000\ \alpha - 2\ 000\ 000)\quad f(\alpha, \beta, \gamma)\, d\alpha\, d\beta\, d\gamma.$$

Soluția optimă este dată de cea mai mare dintre valorile integralelor de mai sus. Astfel dacă prima este cea mai mare, soluția optimă este dată de M (200, 0, 800).

Integralele de mai sus în general nu pot fi caculate direct. În acest caz se folosește metoda numerelor întîmplătoare, pe care am expus-o în unul din capitolele precedente.

2. Programarea stohastică

Exemplul considerat la punctul precedent arată că în foarte multe cazuri elementele care intră în enunțul unei probleme de programare liniară sînt de natură statistică. Ca atare, ele nu pot intra în calcule decît sub forma unor medii, ținînd seamă de distribuția lor statistică. Aceste elemente pot fi deci considerate ca niște variabile aleatoare, a căror lege de probabilitate ne este cunoscută. Pentru acest motiv programarea de acest fel se numește stohastică.

Folosind notația matricială am văzut că o problemă de programare liniară poate fi scrisă sub forma:

$$\begin{aligned} &AX = b, \\ &x \geqslant 0, \\ &cX = \min, \end{aligned} \qquad \text{(XXVI, 2.1)}$$

unde

$$A = (a_{ij}), \qquad (i = 1, 2, \ldots, n;\ j = 1, 2, \ldots, m)$$

$$b = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{pmatrix}, \quad X = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}, \quad c = (c_1, c_2, \ldots, c_n).$$

În cazul de față, funcția obiectiv cX trebuie să fie făcută minimă.

Să considerăm vectorii aleatori

$$\xi = \begin{pmatrix} \xi_1 \\ \xi_2 \\ \vdots \\ \xi_m \end{pmatrix}, \quad \eta = (\eta_1, \eta_2, \ldots, \eta_n) \qquad \text{(XXVI, 2.2)}$$

a căror lege de probabilitate ne este cunoscută şi pentru care avem:

$$\begin{aligned} M(\xi_i) &= b_i, \quad (i = 1, 2, \ldots, m), \\ M(\eta_i) &= c_i, \quad (i = 1, 2, \ldots, n). \end{aligned} \qquad \text{(XXVI, 2.3)}$$

Să considerăm problema de programare:

$$\begin{aligned} AX &= \xi, \\ X &\geqslant 0, \\ \eta X &= \min, \end{aligned} \qquad \text{(XXVI, 2.4)}$$

unde A, X, ξ, η sînt definiţi mai sus.

Relaţiile (XXVI, 2.4) definesc o problemă de programare stohastică.

Mai general, putem considera că şi elemantele matricei A sînt variabile aleatoare. În acest caz, soluţionarea problemei devine foarte dificilă.

La punctul precedent, am tratat o problemă care se enunţă prin relaţiile:

$$\begin{aligned} A\,X &= b, \\ X &\geqslant 0, \\ \eta X &= \max \end{aligned} \qquad \text{(XXVI, 2.5)}$$

în care numai η este un vector aleator.

Am văzut cum se obţine soluţia unei astfel de probleme. Funcţia obiectiv devine maximă într-unul din vîrfurile poliedrului convex. Fie

$$x^{(i)} = \begin{pmatrix} x_1^{(i)} \\ x_2^{(i)}, \\ \vdots \\ x_n^{(i)} \end{pmatrix}, \qquad (i = 1, 2, \ldots, p)$$

unul din vîrfurile acestui poliedru. Cum numărul vîrfurilor este finit, notăm prin p numărul vîrfurilor.

Pentru vîrful $x^{(i)}$, funcţia obiectiv ia valoarea $x^{(i)}$. Pentru ca această valoare să reprezinte soluţia problemei, trebuie să avem:

$$\eta\,(x^{(i)} - x^{(j)}) \geqslant 0, \quad (j = 1, 2, \ldots, p\,;\ j \neq i) \qquad \text{(XXVI, 2.6)}$$

Scanned with OKEN Scanner

Fie R_n spațiul valorilor $\eta = (\eta_1, \eta_2, \ldots, \eta_n)$. În acest spațiu inegalitățile (XXVI, 2.6) determină o regiune T_i, $(i = 1,2,\ldots,p)$.

Este de observat că regiunile T_i și T_j, $(i \neq j)$, nu au puncte comune, deoarece nu putem avea în același timp:

$$\eta(x^{(i)} - x^{(j)}) \geqslant 0, \quad \eta(x^{(j)} - x^{(i)}) \geqslant 0, \quad x^{(i)} \neq x^{(j)}.$$

Să ne alegem un punct η de coordonate $\eta_1, \eta_2, \ldots, \eta_n$ în spațiul T_i. Pentru acest punct valoarea maximă a funcției obiectiv este:

$$\eta x^{(i)} = \eta_1 x_1^{(i)} + \eta_2 x_2^{(i)} + \ldots + \eta_n x_n^{(i)}. \qquad \text{(XXVI, 2.7)}$$

Să notăm prin $F(\eta_1, \eta_2, \ldots, \eta_n)$ funcția de repartiție a variabilei η în spațiul R_n.

Valoarea medie a expresiei (XXVI, 2.7) este

$$S_i = \int\int \ldots \int_{T_i} (\eta_1 x_1^{(i)} + \eta_2 x_2^{(i)} + \ldots + \eta_n x_n^{(i)}) dF(\eta_1 \eta_2 \ldots \eta_n). \text{(XXVI, 2.8)}$$

Deci dacă alegem soluția posibilă $x^{(i)}$, avem valoarea optimă S_i. Este evident că ne alegem atunci soluția $x^{(k)}$ pentru care $S_k \geqslant S_i$, oricare ar fi $i = 1, 2, \ldots, p$; $i \neq k$.

Calculul integralei (XXVI, 2.8) este de cele mai multe ori imposibil de efectuat prin mijloacele obișnuite ale analizei matematice. Trebuie în general să se aplice metode de cvadratură mecanică.

Cum și aceste metode conduc la calcule numeroase, cea mai indicată metodă este aceea a numerelor întîmplătoare, pe care am expus-o la cap. XXIII. Este de reținut, prin urmare, că pentru problema de programare liniară exprimată prin relațiile (XXVI, 2.5) există o metodă generală pentru obținerea soluției.

În cazul cînd legea de probabilitate a vectorului η este discretă, soluția problemei este mai simplă.

Fie p_i probabilitatea egalității $\eta = \eta^i$, $(i = 1,2,\ldots,s)$, admițînd că vectorul η poate lua o valoare η^i $\left(\sum_{i=1}^{s} p_i = 1\right)$.

Cu probabilitatea p_i vom avea deci de rezolvat următoarea problemă de programare liniară:

$$\begin{aligned} &Ax = b, \\ &x \geqslant 0, \\ &\eta^i x = \max. \end{aligned} \qquad \text{(XXVI, 2.9)}$$

Această problemă poate fi rezolvată prin metoda Simplex. Notăm prin x^i soluția optimă și prin L^i valoarea maximă a funcției obiectiv. Soluția optimă corespunde speranței matematice cea mai mare dintre valorile L^i. Trebuie să observăm însă că soluția optimă x^i, care se obține cu probabilitatea p_i, cade, conform teoriei generale, într-unul din vîrfurile poliedrului convex definit prin relațiile:

$$\begin{aligned} &Ax = b, \\ &x \geqslant 0. \end{aligned}$$

Scanned with OKEN Scanner

Acest poliedru rămîne însă același pentru toate sistemele (XXVI, 2.9). Cele s soluții x^i, $(i = 1, 2, \ldots, s)$, ale sistemului (XXVI, 2.9) pot să nu fie distincte între ele. Notăm prin:

$$\eta^{i_1}, \eta^{i_2}, \ldots, \eta^{i_k}$$

vectorii care conduc la soluțiile:

$$x^{i_1} = x^i, \; x^{i_2} = x^i \ldots, \; x^{i_k} = x^i.$$

Admitem deci că un număr k de soluții ale sistemului (XXVI, 2.9) conduc la vectorul x^i, corespunzător unui vîrf al poliedrului convex. Fie

$$L^{i_1}, L^{i_2} \ldots L^{i_k}$$

valorile maxime ale funcției obiectiv corespunzătoare vectorilor x^{i_h}, $(h = 1, 2, \ldots, k)$, în urma rezolvării sistemului (XXVI, 2.9).

Speranța matematică a valorilor L^{i_h}, $(h = 1, 2, \ldots, k)$, este

$$\sum_{h=1}^{k} p_{i_h} L^{i_h}. \qquad \text{(XXVI, 2.10)}$$

Ea ne dă valoarea medie a funcției optime pentru soluțiile care cad în vîrful x^i al poligonului convex.

Dintre toate valorile (XXVI, 2.10) alegem cea mai mare. Vîrful x^i, corespunzător acestei valori, ne dă soluția problemei.

În același mod se rezolvă problema de programare stohastică mai generală:

$$\begin{aligned} Ax &= \xi, \\ x &\geqslant 0, \\ \eta x &= \max, \end{aligned} \qquad \text{(XXVI, 2.11)}$$

unde ξ și η sînt vectorii aleatori

$$\xi = \begin{pmatrix} \xi_1 \\ \xi_2 \\ \vdots \\ \xi_m \end{pmatrix}, \qquad \eta = (\eta_1, \eta_2, \ldots, \eta_n).$$

Fie p_{ij} probabilitatea egalității $\xi = \xi^i$, $\eta = \eta^j$:

$$p_{ij} = P(\xi = \xi^i, \; \eta = \eta^j), \quad (i = 1, 2, \ldots, m; \; j = 1, 2, \ldots, n)$$

Avem

$$p_{ij} \geqslant 0, \; \sum_{i=1}^{m} \sum_{j=1}^{n} p_{ij} = 1.$$

Cu probabilitatea p_{ij} avem de rezolvat problema de programare liniară:

$$Ax = \xi,$$
$$x \geqslant 0, \qquad \text{(XXVI, 2.12)}$$
$$\eta^j x = \max, \ (i=1,2,\ldots,m;\ j=1,2,\ldots,n).$$

Problema se rezolvă prin metodele generale, ca de exemplu metoda Simplex.

Fie x^{ij} soluția corespunzătoare vectorilor ξ^i, η^j și L^{ij} valoarea maximă a funcției optime. Ea are speranța matematică $p_{ij}\ L^{ij}$. Dintre toate valorile $p_{ij}\ L^{ij}$, $(i = 1, 2, \ldots, m;\ j = 1, 2, \ldots, n)$ alegem pe cea mai mare care ne dă astfel soluția optimă.

Este de observat că în cazul de față poliedrul convex

$$Ax = \xi, \qquad \text{(XXVI, 2.13)}$$
$$x \geqslant 0,$$

variază cu $i = 1, 2, \ldots, m$.

Prin urmare, vectori x^{ij} sînt în general diferiți între ei. În cazul cu totul particular, cînd doi dintre acești vectori coincid se aplică metoda arătată în exemplul precedent.

Să considerăm din nou sistemul (XXVI, 2.11). Deoarece ξ și η sînt vectori aleatori, înseamnă că și soluția optimă este o variabilă aleatoare. La fel și valoarea maximă a funcției optime este o variabilă aleatoare. Notăm prin $L(\xi; \eta)$ această valoare maximă, punînd în evidență variabilele ξ și η de care depind. Cunoscînd legile de repartiție ale variabilelor ξ și η din datele problemei de programare, se pune în mod natural studierea variabilei aleatoare $L(\xi; \eta)$. Aceasta este o problemă dificilă, care pînă acum n-a fost soluționată în cazul general. Citățm în această direcție următoarea teoremă datorită lui Vajda.

Să considerăm următoarea problemă de programare stohastică

$$Ax = \xi,$$
$$x \geqslant 0,$$
$$cx = \min.$$

Să notăm prin $L(0)$ valoarea minimă a funcției obiectiv pentru $\xi = 0$. În general, $L(0)$ este o variabilă aleatoare, a cărei valoare medie o notăm prin $M[L(\xi)]$. Să considerăm acum soluția $L(0)$ pentru $0 = M(\xi)$, adică $L(M(\xi)$

S. Vajda a demonstrat că

$$L(M(\xi)) \leqslant M(L(\xi)).$$

Pentru demonstrație trimitem la recentul său tratat asupra programării liniare*).

*) V. Vajda, *Mathematical Programming*, 1961.

În legătură cu studiul soluțiilor aleatoare ale problemelor de programare liniară stohastică, o contribuție importantă au adus recent M. Iosifescu și R. Theodorescu.

Presupunem de data aceasta că problema de programare stohastică se exprimă prin relațiile:

$$\begin{aligned} Ax &= \xi, \\ x &\geqslant 0, \\ \eta x &= \min, \end{aligned} \qquad \text{(XXVI, 2.14)}$$

unde nu numai vectorii ξ și η sînt aleatori, ci și elementele matricei A.

Să notăm prin $M(a_{ij})$ și $D(a_{ij})$ respectiv valoarea medie și dispersia a unui element a_{ij} al lui A. Fie λ un număr real mai mare decît 1.

Introducem matricele:

$$\begin{aligned} A_\lambda^- &= (M(a_{ij}) - \lambda D(a_{ij}), \\ A_\lambda^+ &= (M(a_{ij}) + \lambda D(a_{ij}). \end{aligned}$$

În mod analog, introducem vectorii coloană B_λ^-, B_λ^+ și vectorii linie C_λ^{-1}, C_λ^+ ale căror componente sînt

$$\begin{aligned} B_\lambda^- &= (M(\xi_k) - D(\xi_k), \\ B_\lambda^+ &= (M(\xi_k) + D(\xi_k), \\ C_\lambda^- &= (M(\eta_e) - D(\eta_e), \\ C_\lambda^+ &= (M(\eta_e) + D(\eta_e). \end{aligned}$$

Considerăm acum următoarele două probleme de programare liniară nestohastică:

$$\begin{aligned} A_\lambda^- x &= B_\lambda^- \\ x &\geqslant 0 \\ C_\lambda^- x &= \min \end{aligned} \qquad \text{(XXVI, 2.15)}$$

și

$$\begin{aligned} A_\lambda^+ x &= B_\lambda^+ \\ x &\geqslant 0 \\ C_\lambda^+ x &= \min. \end{aligned} \qquad \text{(XXVI, 2.16)}$$

Fie

$$\begin{aligned} L_\lambda^- &= \min_x C_\lambda^- x, \\ L_\lambda^+ &= \min_x C_\lambda^+ x. \end{aligned}$$

soluțiile optimale ale problemelor (XXVI, 2.15) și (XXVI, 2.16).

Scanned with OKEN Scanner

Se poate arăta că întotdeauna $L_\lambda^- \leqslant L_\lambda^+$ și că poliedrul soluțiilor posibile ale primei probleme este conținut în poliedrul soluțiilor celei de-a doua probleme. Notăm prin Δ_λ regiunea din spațiul m — dimensional egală cu diferența acestor două poliedre.

M. Iosifescu și R. Theodorescu au demonstrat următoarea interesantă teoremă [1]).

„Probabilitatea ca valoarea minimă a funcției obiectiv a programării stohastice (XXVI, 2.14) să fie cuprinsă între L_λ^- și L_λ^+ și ca soluțiile x să aparțină regiunii Δ_λ este cel puțin egală cu $\left(1-\frac{1}{\lambda^2}\right)^{mn+m+n}$".

3. Programarea dinamică

Din expunerea făcută pînă acum asupra programării liniare putem trage următoarele concluzii :

1°. Programarea liniară este în esență de natură statistică. Este evident acest lucru, deoarece în primul rînd, toate datele care se folosesc într-o programare se obțin pe bază de cercetări statistice. Astfel, cînd în problema transporturilor afirmăm că cheltuiala pentru transportul unei unități de marfă de la stația A la depozitul B este suma c lei, ne referim în mod tacit la o medie a cheltuielilor, care este rezultatul unor calcule statistice. Observația este generală. Programarea oricărei acțiuni de rentabilizare reprezintă, în fond, planificarea optimă a unei acțiuni economice, care se va desfășura în viitor. Ca în orice problemă de planificare, ne folosim pentru rezolvarea ei de cunoașterea realității economice , caracterizată în prezent prin anumite date statistice. Analiza economică minuțioasă a acestor date face posibilă previziunea, în anumite limite, a desfășurării viitoare a fenomenului cercetat.

De multe ori într-o problemă de programare liniară nu este însă posibilă exprimarea datelor, de la care pornim prin numere, ci trebuie să considerăm anumite distribuții statistice. După cum am arătat, nu corespunde realității rezolvarea unei probleme de programare liniară în care se afirmă că producția la hectar a unei culturi este de 1 500 kg. În mod corect, planificarea optimă se va obține pornind de la ipoteza că recolta *medie* a culturii este de 1 500 kg și că ea se repartizează în jurul acestei medii după o anumită distribuție statistică, a cărei dispersie o cunoaștem pe baza unor date statistice. În acest caz, programarea liniară va fi nevoită să aplice din plin noțiunile și metodele statisticii matematice. Programarea liniară devine stohastică.

2°. Aplicarea programării liniare în practică depășește cu mult schemele matematice studiate. Ca și în fizică, chimie și în celelalte științe, modelele matematice nu pot cuprinde întreaga realitate, care este cu mult mai complexă și depinde de un număr mult mai mare de factori

[1]) M. Iosifescu și R. Theodorescu, *Asupra programării liniare stohastice*. Analele Universității „C. I. Parhon" (1961).

decît cei folosiţi în calcule. Considerarea numai a unui număr limitat de factori, care sînt cei mai importanţi, şi simplificarea acţiunilor acestor factori, pentru a putea intra într-o formulare cantitativă fac posibilă obţinerea unei scheme matematice, al cărei studiu contribuie, la cunoaşterea realităţii, însă numai cu oarecare aproximaţie. În interpretarea pe care o dă pentru cunoaşterea şi folosirea fenomenelor naturii, omul de ştiinţă ţine seama de faptul mai sus semnalat. La fel şi în economie, cei care au aplicat metoda programării liniare la exemple concrete, s-au lovit de dificultăţi ce nu pot fi traduse într-o formulare matematică. De aceea, programarea liniară nu poate oferi decît o soluţie aproximativă. Rolul prim rămîne tot al economistului care, făcînd o analiză a datelor economice şi a soluţiilor oferite de programarea liniară, va lua măsurile cele mai adecvate pentru obţinerea planului optim. Mai trebuie să ţinem seama şi de faptul că planul se desfăşoară în timp şi că realizarea lui, ţinînd seama de inevitabilele abateri de la previziunile matematice, cade tot în sarcina economistului. Programarea liniară constituie numai unul din elementele de care trebuie să se ţină seama în analiza unei activităţi economice.

3°. Programarea liniară este un caz particular de programare. Ea pune condiţia ca atît funcţia optimă, cît şi relaţiile de legătură să fie liniare. Aceste restricţii în aplicaţiile practice sînt însă numai rareori îndeplinite. De aceea, programarea liniară este destul de rar întîlnită, limitîndu-se la cele cîteva exemple clasice expuse în tratate. Sînt frecvente cazurile cînd funcţia de optimizat nu este liniară. În cazul cînd această funcţie este de forma

$$f(x_1, x_2, \ldots, x_n) = \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij} x_i x_j + \sum_{i=1}^{n} b_i x_i + c \qquad \text{(XXVI, 3.1)}$$

avem de-a face cu o programare pătratică.

Dacă funcţia de optimizat este de forma

$$f(x_1, x_2, \ldots, x_n) = \frac{\sum_{i=1}^{n} a_i x_i + b}{\sum_{i=1}^{n} c_i x_i + d} \qquad \text{(XXVI, 3.2)}$$

programarea se numeşte hiperbolică.

Cînd funcţia de optimizat este un polinom în variabilele $x_1, x_2, \ldots, x_n$, avem o programare polinomială.

Legăturile pentru aceste programări pot, sau nu, să fie liniare.

Studiul general al programărilor este, prin urmare, vast. Trebuie să reţinem faptul important că programarea liniară reprezintă doar un caz particular al unei teorii generale, care interesează pe economişti, statisticieni, matematicieni, ingineri şi pe alţi oameni de ştiinţă aplicată.

Toate exemplele considerate pînă acum se referă la cazul cînd problema ce ni se pune trebuie să fie rezolvată la un anumit moment. Există însă probleme cu caracter economic care se desfăşoară în timp în mai multe etape. În acest caz problema de programare se numeşte dinamică.

Scanned with OKEN Scanner

Diferența esențială dintre o problemă obișnuită de programare și o problemă de programare dinamică este că prima se desfășoară într-o singură etapă și a doua în mai multe etape.

Să admitem că în problema de programare dinamică avem m etape. Putem să alegem un plan de lucru, constînd în deciziile pe care trebuie să le luăm în fiecare etapă. Evident că dintre toate planurile de lucru trebuie să alegem planul optim. Principiul fundamental care stă la baza rezolvării problemelor de programare dinamică, cunoscut sub numele de „principiul optimalității" este următorul: un plan optim pentru m etape se bucură de proprietatea că, oricum ar fi starea inițială și decizia inițială, cele $(m-1)$ decizii care urmează trebuie să constituie un plan optim pe cele $(m-1)$ etape care rămîn după decizia luată în prima etapă.

Cum funcționează acest principiu rezultă din exemplul care urmează.

Într-o gospodărie colectivă se pune problema întocmirii unui plan optim de creștere a unor animale.

Pentru simplificare considerăm o singură categorie de animale, de exemplu porcine, simplificînd datele pentru a nu complica schema matematică. Ne permitem să facem acest lucru pentru motivul că scopul nostru este să rezolvăm o problemă teoretică din care să rezulte principiile și modul de aplicare a programării dinamice.

Să presupunem că inițial gospodăria colectivă posedă porci de trei categorii: n_0 — porci de vîrstă sub 1 an, n_1 — porci de vîrstă cuprinsă între 1 an și 2 ani, n_2 — porci mai mari de 2 ani.

Admitem că un porc de categoria a II-a după un an are în medie k_1 urmași de categoria I (în vîrstă de la 0 la 1 an), iar cei din categoria a III-a au k_2 urmași tot de categoria I.

Dintre acești porci, unii sînt reținuți pentru prăsilă, iar alții vînduți.

Se pune problema întocmirii unui plan de creștere și vînzare a porcilor astfel ca în timp de t ani venitul obținut din vînzarea porcilor să fie maxim, iar numărul porcilor după trecerea acestei perioade de timp să fie cel puțin egal cu m_0, m_1, m_2, respectiv, pentru cele trei categorii.

Notăm prin n_0^s, n_1^s, n_2^s, numărul porcilor existenți la începutul anului s și prin x_0^s, x_1^s, x_2^s, numărul porcilor vînduți la începutul aceluiași an.

Vom avea relațiile de recurență:

$$\begin{aligned} n_0^{s+1} &= k_1(n_1^s - x_1^s) + k_2(n_2^s - x_2^s) \\ n_1^{s+1} &= n_0^s - x_0^s \\ n_2^{s+1} &= n_1^s + n_2^s - x_1^s - x_2^s \end{aligned} \qquad \text{(XXVI, 3.3)}$$

care ne permit din aproape în aproape să determinăm numărul porcilor la începutul fiecărui an. Este evident că trebuie să avem îndeplinite relațiile:

$$0 \leqslant x_0^s \leqslant n_0^s;\ 0 \leqslant x_1^s \leqslant n_1^s;\ 0 \leqslant x_2^s \leqslant n_2^s;\ s = 1, 2, \ldots, t+1, \qquad \text{(XXVI, 3.4)}$$

$$n_0^{t+1} \geqslant m_0;\quad n_1^{t+1} \geqslant m_1;\quad n_2^{t+1} \geqslant m_2. \qquad \text{(XXVI, 3.5)}$$

Relațiile (XXVI, 3.3) (XXVI, 3.4) (XXVI, 3.5) ne dau evoluția șeptelului de animale în perioada de timp considerată. Trebuie să determinăm variabilele x_0^s, x_1^s, x_2^s în așa fel încît venitul rezultat din vînzarea porcilor să fie maxim, fiind îndeplinite condițiile (XXVI, 3.4), (XXVI, 3.5).

Să notăm prin $\varphi_s(x_0^s, x_1^s, x_2^s)$ venitul net rezultat din vînzarea porcilor la începutul anului s; $\varphi_s(x_0^s, x_1^s, x_2^s)$ se obține din suma realizată din vînzare din care am scăzut cheltuielile de întreținere a animalelor pe anul precedent.

Fie $f_i(n_0, n_1, n_2)$ venitul maxim realizat pe o perioadă de i ani, în ipoteza că la începutul perioadei numărul respectiv al porcilor din categoriile I, II, III este n_0, n_1, n_2 și că relațiile (XXVI, 3.4) și (XXVI, 3.5) sînt îndeplinite.

Avem relațiile de recurență:

$$\begin{aligned}
f_t(n_0^1, n_1^1, n_2^1) &= \text{Max}\,[\varphi_1(x_0^1, x_1^1, x_2^1) + f_{t-1}(n_0^2, n_1^2, n_2^2)] \\
f_{t-1}(n_0^2, n_1^2, n_2^2) &= \text{Max}\,[\varphi_2(x_0^2, x_1^2, x_2^2) + f_{t-2}(n_0^3, n_1^3, n_2^3)] \qquad \text{(XXVI, 3.6)} \\
f_{t-2}(n_0^3, n_1^3, n_2^3) &= \text{Max}\,[\varphi_3(x_0^3, x_1^3, x_2^3) + f_{t-3}(n_0^4, n_1^4, n_2^4)] \\
&\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\
f_1(n_0^t, n_1^t, n_2^t) &= \text{Max}\,[\varphi_t(x_0^t, x_1^t, x_2^t)].
\end{aligned}$$

Maximele acestor funcții sînt luate în raport cu variabilele x_0^s, x_1^s, x_2^s, $(s = 1, 2, \ldots, t)$, astfel încît relațiile (XXVI, 3.3) (XXVI, 4.5), (XXVI, 3.5) să fie satisfăcute. Relațiile (XXVI, 3.6) pun în evidență trecerea succesivă de la o perioadă la alta. Procesul economic considerat este cu mai multe stadii, după cum rezultă clar din natura egalităților (XXVI, 3.6).

De cele mai multe ori trecerea de la un stadiu la altul este de natură probabilistică, astfel încît rezolvarea problemei se face cu ajutorul teoriei proceselor stohastice.

Astfel, chiar în cazul simplu pe care l-am examinat, o analiză mai minuțioasă ne obligă să introducem elemente de natură probabilistică. În primul rînd, trebuie să observăm că este mai exact să afirmăm că un animal are un număr k de urmași cu o probabilitate p_k, decît să admitem că are k urmași.

În al doilea rînd, intervine și mortalitatea care poate impune, la un moment dat, tăierea unui mare număr de animale, care de asemenea este de natură probabilistică.

Admițînd că intensitatea înmulțirii porcilor și a mortalității lor dintr-un an este independentă de intensitatea înmulțirii și mortalității din anul precedent, se recunoaște ușor că procesul probabilistic, care intervine în rezolvarea problemei de mai sus, este un proces stohastic cu creșteri independente.

Se pun astfel numeroase probleme ce merită să fie tratate pînă la capăt, atît din punct de vedere matematic cît și practic. Pot fi schimbate, în special, relațiile (XXVI, 3.3) (XXVI, 3.4) (XXVI, 3.5) prin

altele, după aspectul economic al problemei. Aceasta atrage după sine, evident, și schimbarea metodei de soluționare a problemei de maxim.

Rezolvarea problemei din exemplul considerat mai sus depinde de natura funcțiilor $\varphi_s(x_0^1, x_1^1, x_2^1)$.

Vom arăta cum se rezolvă problema în cazul cel mai simplu, cînd prețurile respective pentru un porc din categoria I, II, III sînt constante și egale cu a_0, a_1, a_2, iar cheltuielile anuale de întreținere sînt respectiv b_0, b_1, b_2, și $t = 2$.

În acest caz, avem:

$$\varphi_1(x_0^1, x_1^1, x_2^1) = a_0 x_0^1 + a_1 x_1^1 + a_2 x_2^1,$$

$$\varphi_2(x_0^2, x_1^2, x_2^2) = a_0 x_0^2 + a_1 x_1^2 + a_2 x_2^2 - b_0(n_0^1 - x_0^1) - b_1(n_1^1 - x_1^1) - b_2(n_2^1 - x_2^1).$$

Relațiile (XXVI, 3.3) ne dau succesiv:

$$n_0^1 = n_0;\ n_1^1 = n_1;\ n_2^1 = n_2,$$

$$n_0^2 = k_1(n_1^1 - x_1^1) + k_2(n_2^1 - x_2^1);\ n_1^2 = n_0^1 - x_0^1;\ n_2^2 = n_1^1 + n_2^1 - x_1^1 - x_2^1,$$

$$n_0^3 = k_1(n_0^1 - x_0^1 - x_1^2) + k_2(n_1^1 + n_2^1 - x_1^1 - x_2^1 - x_2^2), \qquad \text{(XXVI, 3.7)}$$

$$n_1^3 = k_1(n_1^1 - x_1^1) + k_2(n_2^1 - x_2^1) - x_0^2,$$

$$n_2^3 = n_0^1 - x_0^1 - x_1^2 + n_1^1 + n_2^1 - x_1^1 - x_2^1 - x_2^2.$$

Trebuie să facem maximă funcția:

$$f(x_0^1, x_1^1, x_2^1, x_0^2, x_1^2, x_2^2) = (a_0 + b_0) x_0^1 + (a_1 + b_1) x_1^1 + (a_2 + b_2) x_2^1 + a_0 x_0^2 + a_1 x_1^2 + a_2 x_2^2 - b_0 n_0 - b_1 n_1 - b_2 n_2 \qquad \text{(XXVI, 3.8)}$$

fiind îndeplinite condițiile:

$$k_1(n_1 - x_1^1) + k_2(n_2 - x_2^1) > 0,$$

$$n_0 - x_0^1 \geqslant 0,$$

$$n_1 + n_2 - x_1^1 - x_2^1 \geqslant 0,$$

$$k_1(n_0 - x_0^1 - x_1^2) + k_2(n_1 + n_2 - x_1^1 - x_2^1 - x_2^2) \geqslant m_0,$$

$$k_1(n_1 - x_1^1) + k_2(n_2 - x_2^1) - x_0^2 \geqslant m_1,$$

$$n_0 + n_1 + n_2 - x_0^1 - x_1^1 - x_1^2 - x_2^1 - x_2^2 \geqslant m_2,$$

$$x_0^1 \geqslant 0,\ x_1^1 \geqslant 0,\ x_2^1 \geqslant 0,\ x_0^2 \geqslant 0,\ x_1^2 \geqslant 0,\ x_2^2 \geqslant 0.$$

Relațiile (XXVI, 3.7) și forma (XXVI, 3.8) a funcției de optimizat ne arată că în acest caz particular se ajunge la o problemă de programare liniară dar numai formal. Aceasta pentru că putem aplica metoda „simplex" pentru determinarea necunoscutelor x_0^2, x_1^2, x_2^2, dar nu le putem folosi pe toate în practică. Problema noastră își păstrează caracterul dinamic, trebuind să fie rezolvată în mai multe etape.

Referindu-ne în legătură cu aceasta tot la cazul general, a cărui soluționare este dată de formulele (XXVI, 3.6), este necesar să facem o

observație esențială, foarte importantă pentru aplicarea programării dinamice. Să admitem că rezolvînd problema determinăm valorile x_0^1, x_1^1, x_2^1, care ne dau respectiv numărul vînzărilor de porci din cele trei categorii în primul an de aplicare a planului. La sfîrșitul primului an, în mod teoretic, ar trebui să avem următoarele numere de animale:

$$n_0^2 = k_1 (n_1 - x_1^1) + k_2(n_2 - x_2^1),$$
$$n_1^2 = n_0 - x_0^1,$$
$$n_2^2 = n_1^1 + n_2^1 - x_1^1 - x_2^1.$$

După trecerea primului an ne vom găsi în realitate în fața altor numere $\overline{n_0^2}, \overline{n_1^2}, \overline{n_2^2}$, în general apropiate de n_0^2, n_1^2, n_2^2, dar diferite de ele.

Aceasta pentru că noi afirmăm că în medie $(n_1 - x_1^1)$ din categoria a doua vor da prin prăsilă $k_1(n_1 - x_1^1)$ urmași și că $(n_2 - x_2^1)$ porci din categoria ultimă vor avea $k_2(n_2 - x_2^1)$ urmași, dar întotdeauna există o diferență între frecvențe și previziunile teoretice. Pentru a lucra în cele mai bune condiții, ținînd seama de realitate, ar trebui să calculăm un nou plan, pe o perioadă de $t-1$ ani, pornind de la numerele inițiale $\overline{n_0^2}$, $\overline{n_1^2}$, $\overline{n_2^2}$. Vom determina astfel numărul vînzărilor din anul al doilea, iar la sfîrșitul anului al doilea vom calcula planul pe o perioadă de $t-2$ ani în funcție de datele reale $\overline{n_0^3}$, $\overline{n_1^3}$, $\overline{n_2^3}$. Pînă la sfîrșit va trebui să întocmim t planuri după metoda indicată.

Din cele de mai sus rezultă că aplicarea programării dinamice se face, pe baza rezultatelor teoretice obținute, numai pentru primul an, deși noi ținem seama în calcule de evoluția procesului pe toate cele t etape. După trecerea fiecărei etape calculele se refac, ajustînd datele obținute prin programare, în baza evoluției reale a procesului economic considerat.

Legătura dintre etape, după cum am mai spus, are forme variate. De exemplu, dacă se pleacă de la cunoscuta schemă matematică a legăturilor dintre ramuri de care ne-am ocupat în capitolul precedent și se pune problema găsirii unui plan optim pentru ca după o perioadă de t ani să se obțină o producție maximă într-una sau mai multe ramuri, se ajunge la maximizarea unor relații clasice care intervin în studiul lanțurilor Markov. În acest caz se aplică, deci, schema lanțurilor lui Markov. În alte cazuri, se aplică scheme de dependență și mai generale, cum sînt, de pildă, schemele proceselor cu legături complete.

Scanned with OKEN Scanner

TABELĂ-ANEXĂ I

$$g(t) = \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} \; ; \; H(t) = \frac{1}{\sqrt{2\pi}} \int_0^t e^{-\frac{x^2}{2}} \, dx.$$

t	$g(t)$	$H(t)$	t	$g(t)$	$H(t)$	t	$g(t)$	$H(t)$
.00	.39894	.00000	.45	.36053	.17364	.90	.26609	.31594
.01	.39892	.00399	.46	.35889	.17724	.91	.26369	.31859
.02	.39886	.00798	.47	.35723	.18082	.92	.26129	.32121
.03	.39876	.01197	.48	.35553	.18439	.93	.25888	.32381
.04	.39862	.01595	.49	.35381	.18793	.94	.25647	.32639
.05	.39844	.01994	.50	.35207	.19146	.95	.25406	.32894
.06	.39822	.02392	.51	.35029	.19497	.96	.25164	.33147
.07	.39797	.02790	.52	.34849	.19487	.97	.24923	.33398
.08	.39767	.03188	.53	.34667	.20194	.98	.24681	.33646
.09	.39733	.03586	.54	.34482	.20540	.99	.24439	.33891
.10	.39695	.03983	.55	.34294	.20884	1.00	.24197	.34134
.11	.39654	.04380	.56	.34105	.21226	1.01	.23955	.34375
.12	.39608	.04776	.57	.33912	.21566	1.02	.23713	.34614
.13	.39559	.05172	.58	.33718	.21904	1.03	.23471	.34850
.14	.39505	.05567	.59	.33521	.22240	1.04	.23230	.35083
.15	.39448	.05962	.60	.33322	.22575	1.05	.22988	.35314
.16	.39387	.06356	.61	.33121	.22907	1.06	.22747	.35543
.17	.39322	.06749	.62	.32918	.23237	1.07	.22506	.35769
.18	.39253	.07142	.63	.32713	.23565	1.08	.22265	.35993
.19	.39181	.07535	.64	.32506	.23891	1.09	.22025	.36214
.20	.39104	.07926	.65	.32297	.24215	1.10	.21785	.36433
.21	.39024	.08317	.66	.32086	.24537	1.11	.21546	.36650
.22	.38940	.08706	.67	.31874	.24857	1.12	.21307	.36864
.23	.38853	.09095	.68	.31659	.25175	1.13	.21069	.37076
.24	.38762	.09483	.69	.31443	.25490	1.14	.20831	.37286
.25	.38667	.09871	.70	.31225	.25804	1.15	.20594	.37493
.26	.38568	.10257	.71	.31006	.26115	1.16	.20357	.37698
.27	.38466	.10642	.72	.30785	.26424	1.17	.20121	.37900
.28	.38361	.11026	.73	.30563	.26730	1.18	.19886	.38100
.29	.38251	.11409	.74	.30339	.27035	1.19	.19652	.38298
.30	.38139	.11791	.75	.30114	.27337	1.20	.19419	.38493
.31	.38023	.12172	.76	.29887	.27637	1.21	.19186	.38686
.32	.37903	.12552	.77	.29659	.27935	1.22	.18954	.38877
.33	.37780	.12930	.78	.29431	.28230	1.23	.18724	.39065
.34	.37654	.13307	.79	.29200	.28524	1.24	.18494	.39251
.35	.37524	.13683	.80	.28969	.28814	1.25	.18265	.39435
.36	.37391	.14058	.81	.28737	.29103	1.26	.18037	.39617
.37	.37255	.14431	.82	.28504	.29389	1.27	.17810	.39796
.38	.37115	.14803	.83	.28269	.29673	1.28	.17585	.39973
.39	.36973	.15173	.84	.28034	.29955	1.29	.17360	.40147
.40	.36827	.15542	.85	.27798	.30234	1.30	.17137	.40320
.41	.36678	.15910	.86	.27562	.30511	1.31	.16915	.40490
.42	.36526	.16276	.87	.27324	.30785	1.32	.16694	.40658
.43	.36371	.16640	.88	.27086	.31057	1.33	.16474	.40824
.44	.36213	.17003	.89	.26848	.31327	1.34	.16256	.40988

Scanned with OKEN Scanner

(continuare)

t	$g(t)$	$H(t)$	t	$g(t)$	$H(t)$	t	$g(t)$	$H(t)$
1.35	.16038	.41149	1.80	.07895	.46407	2.25	.03174	.48778
1.36	.15822	.41309	1.81	.07754	.46485	2.26	.03103	.48809
1.37	.15608	.41466	1.82	.07614	.46562	2.27	.03034	.48840
1.38	.15395	.41621	1.83	.07477	.46638	2.28	.02965	.48870
1.39	.15183	.41774	1.84	.07341	.46712	2.29	.02898	.48899
1.40	.14973	.41924	1.85	.07206	.46784	2.30	.02833	.48928
1.41	.14764	.42073	1.86	.07074	.46856	2.31	.02768	.48956
1.42	.14556	.42220	1.87	.06943	.46926	2.32	.02705	.48983
1.43	.14350	.42364	1.88	.06814	.46995	2.33	.02643	.49010
1.44	.14146	.42507	1.89	.06687	.47062	2.34	.02582	.49036
1.45	.13943	.42647	1.90	.06562	.47128	2.35	.02522	.49061
1.46	.13742	.42786	1.91	.06439	.47193	2.36	.02463	.49086
1.47	.13542	.42922	1.92	.06316	.47257	2.37	.02406	.49111
1.48	.13344	.43056	1.93	.06195	.47320	2.38	.02349	.49134
1.49	.13147	.43189	1.94	.06077	.47381	2.39	.02294	.49158
1.50	.12952	.43319	1.95	.05959	.47441	2.40	.02239	.49180
1.51	.12758	.43448	1.96	.05844	.47500	2.41	.02186	.49202
1.52	.12566	.43574	1.97	.05730	.47558	2.42	.02134	.49224
1.53	.12376	.43699	1.98	.05618	.47615	2.43	.02083	.49245
1.54	.12188	.43822	1.99	.05508	.47670	2.44	.02033	.49266
1.55	.12001	.43943	2.00	.05399	.47725	2.45	.01984	.49286
1.56	.11816	.44062	2.01	.05292	.47778	2.46	.01936	.49305
1.57	.11632	.44179	2.02	.05186	.47831	2.47	.01889	.49324
1.58	.11450	.44295	2.03	.05082	.47882	2.48	.01842	.49343
1.59	.11270	.44408	2.04	.04980	.47932	2.49	.01797	.49361
1.60	.11092	.44520	2.05	.04879	.47982	2.50	.01753	.49379
1.61	.10915	.44630	2.06	.04780	.48030	2.51	.01709	.49396
1.62	.10741	.44738	2.07	.04682	.48077	2.52	.01667	.49413
1.63	.10567	.44845	2.08	.04586	.48124	2.53	.01625	.49430
1.64	.10396	.44950	2.09	.04491	.48169	2.54	.01585	.49446
1.65	.10226	.45053	2.10	.04398	.48214	2.55	.01545	.49461
1.66	.10059	.45154	2.11	.04307	.48257	2.56	.01506	.49477
1.67	.09893	.45254	2.12	.04217	.48300	2.57	.01468	.49492
1.68	.09728	.45352	2.13	.04128	.48341	2.58	.01431	.49506
1.69	.09566	.45449	2.14	.04041	.48382	2.59	.01394	.49520
1.70	.09405	.45543	2.15	.03955	.48422	2.60	.01358	.49534
1.71	.09246	.45637	2.16	.03871	.48461	2.61	.01323	.49547
1.72	.09089	.45728	2.17	.03788	.48500	2.62	.01289	.49560
1.73	.08933	.45818	2.18	.03706	.48537	2.63	.01256	.49573
1.74	.08780	.45907	2.19	.03626	.48574	2.64	.01223	.49585
1.75	.08628	.45994	2.20	.03547	.48610	2.65	.01191	.49598
1.76	.08478	.46080	2.21	.03470	.48645	2.66	.01160	.49609
1.77	.08329	.46164	2.22	.03394	.48679	2.67	.01130	.49621
1.78	.08183	.46246	2.23	.03319	.48713	2.68	.01100	.49632
1.79	.08038	.46327	2.24	.03246	.48745	2.69	.01071	.49643

Scanned with OKEN Scanner

(continuare)

t	$g(t)$	$H(t)$	t	$g(t)$	$H(t)$	t	$g(t)$	$H(t)$
2.70	.01042	.49653	3.15	.00279	.49918	3.60	.00061	4.9984
2.71	.01014	.49664	3.16	.00271	.49921	3.61	.00059	.49985
2.72	.00987	.49674	3.17	.00262	.49924	3.62	.00057	.49985
2.73	.00961	.49683	3.18	.00254	.49926	3.63	.00055	.49986
2.74	.00935	.49693	3.19	.00246	.49929	3.64	.00053	.49986
2.75	.00909	.49702	3.20	.00238	.49931	3.65	.00051	.49987
2.76	.00885	.49711	3.21	.00231	.49934	3.66	.00049	.49987
2.77	.00861	.49720	3.22	.00224	.49936	3.67	.00047	.49988
2.78	.00837	.49728	3.23	.00216	.49938	3.68	.00046	.49988
2.79	.00814	.49736	3.24	.00210	.49940	3.69	.00044	.49989
2.80	.00792	.49744	3.25	.00203	.49942	3.70	.00042	.49989
2.81	.00770	.49752	3.26	.00203	.49944	3.71	.00041	.49990
2.82	.00748	.49760	3.27	.00196	.49946	3.72	.00039	.49990
2.83	.00727	.49767	3.28	.00184	.49948	3.73	.00038	.49990
2.84	.00707	.49774	3.29	.00178	.49950	3.74	.00037	.49991
2.85	.00687	.49781	3.30	.00172	.49952	3.75	.00035	.49991
2.86	.00668	.49788	3.31	.00167	.49953	3.76	.00034	.49992
2.87	.00649	.49795	3.32	.00161	.49955	3.77	.00033	.49992
2.88	.00631	.49801	3.33	.00156	.49957	3.78	.00031	.49992
2.89	.00613	.49807	3.34	.00151	.49958	3.79	.00030	.49992
2.90	.00595	.49813	3.35	.00146	.49960	3.80	.00029	.49993
2.91	.00578	.49819	3.36	.00141	.49961	3.81	.00028	.49993
2.92	.00562	.49825	3.37	.00136	.49962	3.82	.00027	.49993
2.93	.00545	.49831	3.38	.00132	.49964	3.83	.00026	.49994
2.94	.00530	.49836	3.39	.00127	.49965	3.84	.00025	.49994
2.95	.00514	.49841	3.40	.00123	.49966	3.85	.00024	.49994
2.96	.00499	.49846	3.41	.00119	.49968	3.86	.00023	.49994
2.97	.00485	.49851	3.42	.00115	.49969	3.87	.00022	.49995
2.98	.00471	.49856	3.43	.00111	.49970	3.88	.00021	.59995
2.99	.00457	.49861	3.44	.00107	.49971	3.89	.00021	.49995
3.00	.00443	.49865	3.45	.00104	.49972	3.90	.00020	.49995
3.01	.00430	.49869	3.46	.00100	.49973	3.91	.00019	.49995
3.02	.00417	.49874	3.47	.00097	.49974	3.92	.00018	.49996
3.03	.00405	.49878	3.48	.00094	.49975	3.93	.00018	.49996
3.04	.00393	.49882	3.49	.00090	.49976	3.94	.00017	.49996
3.05	.00381	.49886	3.50	.00087	.49977	3.95	.00016	.49996
3.06	.00370	.49889	3.51	.00084	.49978	3.96	.00016	.49996
3.07	.00358	.49893	3.52	.00081	.49978	3.97	.00015	.49996
3.08	.00348	.49897	3.53	.00079	.49979	3.98	.00014	.49997
3.09	.00337	.49900	3.54	.00076	.49980	3.99	.00014	.49997
3.10	.00327	.49903	3.55	.00073	.49981			
3.11	.00317	.49906	3.56	.00071	.49981			
3.12	.00307	.49910	3.57	.00068	.49982			
3.13	.00298	.49913	3.58	.00066	.49983			
3.14	.00288	.49916	3.59	.00063	.49983			

Scanned with OKEN Scanner

TABELĂ-ANEXĂ II

$$P = \frac{1}{2^{n/2}\,\Gamma\left(\frac{n}{2}\right)} \int_{\chi^2}^{\infty} x^{\frac{n}{2}-1} e^{-\frac{x}{2}}\, dx; \quad 100P = p.$$

n \ p	Valorile lui χ^2_p în funcție de n și P									
	$p = 99$	95	90	70	50	30	10	5	1	0.1
1	0.000	0.004	0.016	0.148	0.455	1.074	2.706	3.841	6.635	10.827
2	0.020	0.103	0.211	0.713	1.386	2.408	4.605	5.991	9.210	13.815
3	0.115	0.352	0.584	1.424	2.366	3.665	6.251	7.815	11.341	16.268
4	0.297	0.711	1.064	2.195	3.357	4.878	7.779	9.488	13.277	18.465
5	0.554	1.145	1.610	3.000	4.351	6.064	9.236	11.070	15.086	20.517
6	0.872	1.635	2.204	3.828	5.348	7.231	10.645	12.592	16.812	22.457
7	1.239	2.167	2.833	4.671	6.346	8.383	12.017	14.067	18.475	24.322
8	1.646	2.733	3.490	5.527	7.344	9.524	13.362	15.507	20.090	26.125
9	2.088	3.325	4.168	6.393	8.343	10.656	14.684	16.919	21.666	27.877
10	2.558	3.940	4.865	7.267	9.342	11.781	15.987	18.307	23.209	29.588
11	3.053	4.575	5.578	8.148	10.341	12.899	17.275	19.675	24.725	31.264
12	3.571	5.226	6.304	9.034	11.340	14.011	18.549	21.026	26.217	32.909
13	4.107	5.892	7.042	9.926	12.340	15.119	19.812	22.362	27.688	34.528
14	4.660	6.571	7.790	10.821	13.339	16.222	21.064	23.685	29.141	36.123
15	5.229	7.261	8.547	11.721	14.339	17.322	22.307	24.996	30.578	37.697
16	5.812	7.962	9.312	12.624	15.338	18.418	23.542	24.296	32.000	39.252
17	6.408	8.672	10.085	13.531	16.338	19.511	24.769	27.587	33.409	40.790
18	7.015	9.390	10.865	14.440	17.338	20.601	25.989	28.869	34.805	42.312
19	7.633	10.117	11.651	15.352	18.338	21.689	27.204	30.144	36.191	43.820
20	8.260	10.851	12.443	16.266	19.337	22.775	28.412	31.410	37.566	45.315
21	8.897	11.591	13.240	17.182	20.337	23.858	29.615	32.671	38.932	46.797
22	9.542	12.338	14.041	18.101	21.337	24.939	30.813	33.924	40.289	48.268
23	10.196	13.091	14.848	19.021	22.337	26.018	32.007	35.172	41.638	49.728
24	10.856	13.848	15.659	19.943	23.337	27.096	33.196	36.415	42.980	51.179
25	11.524	14.611	16.473	20.867	24.337	28.172	34.382	37.652	44.314	52.620
26	12.198	15.379	17.292	21.792	25.336	29.246	35.563	38.885	45.642	54.052
27	12.879	16.151	18.114	22.719	26.336	30.319	36.741	40.113	46.963	55.476
28	13.565	16.928	18.939	23.647	27.336	31.391	37.916	41.337	48.278	56.893
29	14.256	17.708	19.768	24.577	28.336	32.461	39.087	42.557	49.588	58.302
30	14.953	18.493	20.599	25.508	29.336	33.530	40.256	43.773	50.892	59.703

Scanned with OKEN Scanner

TABELĂ-ANEXĂ III

$$P = \frac{2}{\sqrt{n\pi}} \frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)} \int_{t_p}^{\infty} \left(1 + \frac{x^2}{n}\right)^{-\frac{n+1}{2}} dx; \quad 100P = p.$$

n \ p	Valorile lui t_p în funcție de n și p							
	$P = 90$	70	50	30	10	5	1	0.1
1	0.158	0.510	1.000	1.963	6.314	12.706	63.657	636.619
2	0.142	0.445	0.816	1.386	2.920	4.303	9.925	31.598
3	0.137	0.424	0.765	1.250	2.353	3.182	5.841	12.941
4	0.134	0.414	0.741	1.190	2.132	2.776	4.604	8.610
5	0.132	0.408	0.727	1.156	2.015	2.571	4.032	6.859
6	0.131	0.404	0.718	1.134	1.943	2.447	3.707	5.959
7	0.130	0.402	0.711	1.119	1.895	2.365	3.499	5.405
8	0.130	0.399	0.706	1.108	1.860	2.306	3.355	5.041
9	0.129	0.398	0.703	1.100	1.833	2.262	3.250	4.781
10	0.129	0.397	0.700	1.093	1.812	2.228	3.169	4.587
11	0.129	0.396	0.697	1.088	1.796	2.201	3.106	4.437
12	0.128	0.395	0.695	1.083	1.782	2.179	3.055	4.318
13	0.128	0.394	0.694	1.079	1.771	2.160	3.012	4.221
14	0.128	0.393	0.692	1.076	1.761	2.145	2.977	4.140
15	0.128	0.393	0.691	1.074	1.753	2.131	2.947	4.073
16	0.128	0.392	0.690	1.071	1.746	2.120	2.921	4.015
17	0.128	0.392	0.689	1.069	1.740	2.110	2.898	3.965
18	0.127	0.392	0.688	1.067	1.734	2.101	2.878	3.922
19	0.127	0.391	0.688	1.066	1.729	2.093	2.861	3.883
20	0.127	0.391	0.687	1.064	1.725	2.086	2.845	3.850
21	0.127	0.391	0.686	1.063	1.721	2.080	2.831	3.819
22	0.127	0.390	0.686	1.061	1.717	2.074	2.819	3.792
23	0.127	0.390	0.685	1.060	1.714	2.069	2.807	3.767
24	0.127	0.390	0.685	1.059	1.711	2.064	2.797	3.745
25	0.127	0.390	0.684	1.058	1.708	2.060	2.787	3.725
26	0.127	0.390	0.684	1.058	1.706	2.056	2.779	3.707
27	0.127	0.389	0.684	1.057	1.703	2.052	2.771	3.690
28	0.127	0.389	0.683	1.056	1.701	2.048	2.763	3.674
29	0.127	0.389	0.683	1.055	1.699	2.045	2.756	3.659
30	0.127	0.389	0.683	1.055	1.697	2.042	2.750	3.646
40	0.126	0.388	0.681	1.050	1.684	2.021	2.704	3.551
60	0.126	0.387	0.679	1.046	1.671	2.000	2.660	3.460
120	0.126	0.386	0.677	1.041	1.658	1.980	2.617	3.373
	0.126	0.385	0.674	1.036	1.645	1.960	2.576	3.291

Scanned with OKEN Scanner

Redactor responsabil: MOCANU P.

Dat la cules 26.01.1962. Bun de tipar 11.09.1962. Apărut 1962. Tiraj 1200 ex. 51,18 Hîrtie velină de 80 g/m² 700×1000. Coli editoriale legate Coli de tipar 45,75. A 04060/1961. C. Z. pentru bibliotecile mari 519.2, C. Z. pentru bibliotecile mici 519.

Intreprinderea poligrafică nr. 2, str. Brezoianu nr. 23 – 25. București, R.P.R. Comanda nr. 560

Scanned with OKEN Scanner